



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS DEPARTAMENTO DE QUÍMICA

Una exploración a las relaciones cuantitativas entre datos espectrométricos y actividad a través de descriptores mecano-cuánticos

TESISTA: HERNÁN RUBÉN SÁNCHEZ
DIRECTOR: REINALDO PIS DIEZ
CODIRECTOR: ANDREW MERCADER

2017

*Esta Tesis Doctoral fue realizada en el Centro de Química
Inorgánica (UNLP-CONICET), entre mayo de 2014 y marzo de
2017, bajo la dirección del Dr. Reinaldo Pis Diez y la codirección
del Dr. Andrew Mercader.*

*Expreso aquí mis agradecimientos a quienes aportaron
científicamente a este trabajo.*

*Agradezco a mis directores; al Dr. Reinaldo Pis Diez por la
libertad y confianza brindadas, al Dr. Andrew Mercader por su
apoyo y disposición constante.*

*A mi compañero Julián Del Pla, por compartir tantas horas de
discusiones enriquecedoras sobre nuestras tesis.*

Índice general

I	INTRODUCCIÓN	1
1.	Introducción	3
1.1.	Objetivo general	3
1.2.	Estructura	4
II	CALCULO DE PROPIEDADES MAGNÉTICAS DE ÁTOMOS Y MOLÉCULAS	7
2.	Hamiltoniano de espín	9
2.1.	Breve resumen de mecánica clásica	9
2.1.1.	Mecánica Newtoniana	9
2.1.2.	Campos electromagnéticos	10
2.1.3.	Mecánica lagrangiana	11
2.1.4.	Fuerzas dependientes de la velocidad	12
2.1.5.	Lagrangiano de una partícula en un campo magnético	13
2.1.6.	Mecánica hamiltoniana	13
2.1.7.	Transformaciones de los potenciales	14
2.1.8.	Transformaciones en ecuaciones de movimiento	14
2.2.	Mecánica relativista	15
2.2.1.	Ecuación de Dirac	17
2.3.	Ecuación de Pauli	20
2.4.	Espín	21
2.5.	Ideas principales sobre el capítulo	23
3.	Propiedades magnéticas	25
3.1.	Hamiltoniano de espín efectivo	25

3.1.1.	Descripción fenomenológica	26
3.2.	Parámetros del hamiltoniano efectivo de espín y derivadas de la energía	31
3.3.	Parámetros de RMN para funciones de onda exactas	31
3.3.1.	Hamiltoniano molecular y forma explícita del campo magnético	32
3.3.2.	Expresiones explícitas de propiedades moleculares	33
3.4.	Cálculo de las propiedades magnéticas	35
3.4.1.	Principios generales para el cálculo en funciones de onda aproximadas	35
3.4.2.	Hamiltoniano electrónico en segunda cuantización	37
3.5.	El problema del origen para el potencial vector magnético	42
3.6.	Ideas principales sobre el capítulo	45
4.	Teoría del Funcional de la Densidad	47
4.1.	Introducción	47
4.2.	Forma explícita de las funcionales	50
4.2.1.	Modelo de Thomas-Fermi-Dirac	50
4.3.	Teoría de Kohn-Sham	52
4.4.	B3LYP	53
4.5.	Cálculo de propiedades magnéticas con DFT	55
4.5.1.	Viabilidad y conveniencia	56
4.6.	Teoría del funcional de la densidad y la corriente	57
4.7.	Ideas principales sobre el capítulo	59
III	TEORÍA QSPR	61
5.	QSPR	63
5.1.	Introducción	63
5.2.	Descriptores	66
5.2.1.	Descriptores base	67
5.3.	Construcción de modelos	69
5.3.1.	Algoritmos de selección de descriptores	73
5.4.	Prueba de modelos	75
5.4.1.	Dominio de aplicabilidad	77

6. Algoritmos desarrollados para la construcción de modelos QSPR	81
6.1. Variación sobre el método reemplazo modificado	81
6.1.1. Mínimos cuadrados, proyecciones y el método de reemplazo . . .	81
6.1.2. Variantes consideradas	84
 IV METODOLOGÍA	 87
 7. Bases de datos	 89
 8. Aspectos prácticos del cálculo de los descriptores	 105
8.1. Introducción	105
8.2. Descripción de la metodología empleada	106
8.2.1. Paso 1: Obtención de geometrías de partida	106
8.2.2. Paso 2: Viabilidad de las geometrías de partida	107
8.2.3. Paso 3: Primera optimización de geometría	107
8.2.4. Paso 4: Estudio conformacional	108
8.2.5. Paso 5: Cálculo de las propiedades magnéticas	111
8.2.6. Cálculo de los descriptores	112
 V RESULTADOS	 115
 9. Viabilidad de los descriptores	 117
9.1. Modelo para predicción de Log P	121
 VI APÉNDICES	 129
 A. Segunda cuantización	 131
A.1. Estructura formal	131
A.2. Aplicación en química computacional	133
 B. Códigos de fuente	 137

Parte I

INTRODUCCIÓN

Capítulo 1

Introducción

En este capítulo se reseña el objetivo y la estructura de esta tesis.

1.1. Objetivo general

Cuando la materia es sometida a un campo magnético uniforme, se modifica la forma en la cual la misma absorbe radiación electromagnética. Dicho cambio depende del campo empleado y de la naturaleza de la materia que es sometida a este.

Una parte útil del espectro de absorción, en estas condiciones, es medido en los experimentos de Resonancia Magnética Nuclear (RMN). A partir de estos espectros puede obtenerse información estructural de las moléculas estudiadas.

En la práctica, el espectro puede ser vinculado con la estructura mediante el análisis de parámetros que son obtenidos en base al espectro. Esto permite disminuir la complejidad de dicho vínculo.

Para esto se utiliza una representación simplificada de las interacciones que tienen lugar en la molécula. La energía correspondiente a este esquema simplificado es explicitada en un hamiltoniano efectivo¹. Este contiene una serie de parámetros que son ajustados al experimento en cuestión. Los mismos pueden ser utilizados luego para análisis estructurales.

Las diferencias entre los espectros de absorción, que corresponden a la presencia o no del campo magnético externo, pueden ser interpretadas a partir de construcciones teóricas basadas en primeros principios. Las absorciones pertinentes son entendidas

¹En este contexto “efectivo” hace referencia a que sólo se contemplan aquellas interacciones de interés para el caso.

como transiciones entre estados correspondientes a distintos niveles de energía, que se diferencian en sus estados de espín nuclear.

El cálculo de los parámetros de RMN no es inmediato, ya que estos están ausentes del hamiltoniano construido a partir de primeros principios. Sin embargo, se los puede estimar a partir de transformaciones de dicho hamiltoniano.

Por conveniencia operativa e interpretativa, los operadores involucrados en estas transformaciones son separados en sumandos, de acuerdo con el tipo de interacción que representan. Los parámetros del hamiltoniano de espín efectivo pueden obtenerse como la suma de aportes individuales de cada término. Estas contribuciones tienen relación con las características estructurales de las especies químicas.

En esta tesis se pretende establecer la factibilidad del uso de estas relaciones en modelos QSPR para la predicción de propiedades fisicoquímicas de interés.

1.2. Estructura

Esta tesis está dividida en Partes. Cada parte contiene al menos un Capítulo. Las Partes II y III son esencialmente teóricas. Cada uno de sus capítulos contienen en su final un resumen. En la parte II se describe la naturaleza de propiedades magnéticas y el cálculo de las mismas. Estas serán relacionadas con otras propiedades fisicoquímicas de interés mediante la teoría QSPR (descrita en Parte III).

Las Partes teóricas se escribieron de modo que puedan ser comprendidas en su mayoría utilizando los conocimientos adquiridos en la Licenciatura en Química de esta Facultad, y que junto a ellos resulten al menos casi auto-contenidas.

En la Parte II se describirá con cierto grado de detalle la construcción del Hamiltoniano de primeros principios (Capítulo 2). Esto es esencial al entendimiento del problema. Las transformaciones que permiten obtener los parámetros de RMN a partir de este se describen en el Capítulo 3. Una vez obtenidas las expresiones buscadas debe resolverse el problema de estructura electrónica. Esto se hizo mediante el uso de la Teoría del Funcional de la Densidad. Se discute brevemente esta teoría y la validez de su uso para tal fin en el Capítulo 4.

La Parte III versa sobre la teoría QSPR. En el Capítulo 5 se describe sus fundamentos y algunas consideraciones prácticas tenidas en cuenta en esta tesis. En el Capítulo 6 se analiza un método de selección de variables usado en este trabajo. También se incluye resultados de variaciones propuestas.

La Parte IV describe aspectos metodológicos. Comprende al Capítulo 7 en el que se

describe las bases de datos utilizadas. También al Capítulo 8 en el que se detalla la forma en que se hicieron los cálculos.

La Parte V contiene al Capítulo 9, en el cual se presentan y discuten los resultados obtenidos.

La Parte VI comprende dos apéndices, en el Apéndice A se describe el formalismo de segunda cuantización, el cual es utilizado en el Capítulo 3. El Apéndice B contiene algunos códigos de fuente creados para esta tesis.

Parte II

CALCULO DE PROPIEDADES MAGNÉTICAS DE ÁTOMOS Y MOLÉCULAS

Capítulo 2

Hamiltoniano de espín

2.1. Breve resumen de mecánica clásica

2.1.1. Mecánica Newtoniana

Una vez establecido un sistema de referencia, la posición de una partícula puntal¹ al tiempo t ($\mathbf{r}(t)$) puede conocerse si se conoce su aceleración $\mathbf{a}(t)$, su posición $\mathbf{r}(t_0)$ y velocidad ($\mathbf{v}(t_0)$) para un tiempo t_0 dado. Se afirma con carácter de ley física la *existencia* de sistemas de referencia inerciales, en estos la aceleración no nula de la partícula se debe a, al menos, una interacción de un agente externo con la partícula. Si dos partículas interactúan entre sí, y no con otros cuerpos, entonces sus aceleraciones referidas a un sistema de referencia inercial tendrán sentidos opuestos y el cociente de sus módulos será una constante. La última afirmación describe una ley física. Dicha constante puede ser expresada como un cociente entre una magnitud propia de cada cuerpo, llamada *masa* (m), cuyo valor para un cuerpo de referencia puede ser establecido arbitrariamente.

Se otorga realidad física a la interacción y su intensidad es cuantificada a través de la *fuerza* \mathbf{F} , que para una dada interacción es *definida*² por

$$\mathbf{F}(\mathbf{r}, \mathbf{v}, t) = m \mathbf{a}$$

Si sobre un cuerpo existe más de una interacción, la fuerza es definida como la superposición aditiva de las fuerzas correspondientes a cada interacción aislada. Esto constituye

¹En este contexto, por simplicidad, se utilizará el término “partícula puntual” para referirse a un cuerpo cuya posición pueda ser modelada como tal.

²Con lo cual se resalta que, en el contexto de este tratamiento, no es una ley física.

otra ley física.

La magnitud de una clase en particular de interacciones, llamadas gravitatorias, es expresada utilizando a la masa como única propiedad intrínseca de sendos cuerpos³.

2.1.2. Campos electromagnéticos

Existe una clase de interacciones (llamadas electromagnéticas) que no puede ser atribuida a la masa, la propiedad de las partículas que las hace susceptibles a esta clase de interacciones se llama electricidad, y se cuantifica con la magnitud *carga eléctrica* (q). Una forma conveniente de expresar la fuerza correspondiente a esta clase de interacciones sobre una partícula con carga q , es mediante la *fuerza de Lorentz*

$$\mathbf{F} = q (\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

donde $\mathbf{E}(\mathbf{r}, t)$ es llamado *campo eléctrico* y $\mathbf{B}(\mathbf{r}, t)$ *campo magnético*, y satisfacen las *ecuaciones de Maxwell*, las cuales en vacío son:

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \times \mathbf{B} - \epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t} &= \mu_0 \mathbf{j}\end{aligned}$$

En ellas ϵ_0 y μ_0 representan a las permitividades eléctrica y magnética de vacío, y ρ y \mathbf{j} a las densidades de carga y corriente.

Todo campo vectorial en $\mathbb{S} \subseteq \mathbb{R}^3$ de clase C^2 cuya divergencia sea cero (campo solenoidal) puede expresarse como el rotor de alguna función potencial, por lo que puede definirse al *vector potencial magnético* (\mathbf{A}) como cualquier campo que cumpla

$$\mathbf{B} = \nabla \times \mathbf{A}$$

³La interpretación más habitual es que estas interacciones existen debido a que los cuerpos tienen la propiedad de tener masa. Aunque en este contexto sea arbitrario considerar más fundamental a la propiedad que a la interacción, por simpleza se utilizará este esquema.

Comentario 2.1.1. Dado que el campo magnético es solenoidal, es posible encontrar un vector potencial haciendo

$$\mathbf{A}(\mathbf{r}) = \int_0^1 dt \mathbf{B}(t\mathbf{r}) \times (\mathbf{r}t)$$

Como el campo eléctrico no es solenoidal, no puede expresárselo en forma análoga. Reemplazando según la ecuación anterior en la ley de Faraday⁴ y reordenando

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0$$

y como el rotacional del campo entre paréntesis es cero, entonces dicho campo debe ser igual al gradiente de una función escalar, a la que se le llama *potencial eléctrico* (ϕ), por lo que puede expresarse al campo eléctrico en función de los potenciales como

$$\mathbf{E} = -\nabla \phi - \frac{\partial \mathbf{A}}{\partial t}$$

2.1.3. Mecánica lagrangiana

La trayectoria de un conjunto de partículas también puede conocerse si se conoce las posiciones iniciales y finales y sendos tiempos, junto a las formas de los potenciales del sistema. Dado un conjunto de n coordenadas de posición generalizadas $\{\mathbf{q}\}$, y las correspondientes coordenadas de velocidad generalizadas $\{\dot{\mathbf{q}}\}$, la trayectoria del sistema entre los tiempos t_1 y t_2 es aquella que haga estacionaria a la *acción* S , (*principio de mínima acción*)

$$S := \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}, t) dt$$

⁵La función L es llamada *lagrangiano*, no existe un procedimiento general para la obtención de la misma. Mediante cálculo de variaciones puede demostrarse que la condición de S estacionaria implica que para cada coordenada q_i se cumple la *ecuación de movimiento*

$$\frac{\partial L}{\partial q_i} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) = 0 \quad (2.1)$$

⁴La ecuación de Maxwell escrita aquí en tercer lugar.

⁵El símbolo $:=$ indica que aquello a su izquierda es definido de modo que signifique aquello que está a su derecha.

Se define al *momento generalizado*⁶ como el vector cuyas componentes i satisfacen

$$p_i(\mathbf{q}, \dot{\mathbf{q}}, t) := \frac{\partial L}{\partial \dot{q}_i} \quad (2.2)$$

que es una constante de movimiento cuando el lagrangiano no depende explícitamente de la posición.

2.1.4. Fuerzas dependientes de la velocidad

Si la fuerza no depende explícitamente del tiempo ni de la velocidad, entonces existe una función potencial V tal que

$$\mathbf{F}(\mathbf{r}) = -\nabla V(\mathbf{r})$$

En tal caso el $\mathbf{F}(\mathbf{r})$ representa un *campo de fuerza*⁷, y dado que la energía es conservada se dice de la fuerza que es *conservativa*, por lo que si T representa la energía cinética de la partícula

$$d(T + V) = 0$$

dado que para cada dirección: $d(T + V) = m\dot{x}\ddot{x}dt - Fdx = (m\ddot{x})(\dot{x}dt) - Fdx = 0$

Condiciones equivalentes a la existencia de V para que $\mathbf{F}(\mathbf{r})$ sea conservativa son, que el dominio de $\mathbf{F}(\mathbf{r})$ sea simplemente conexo y $\nabla \times \mathbf{F}(\mathbf{r}) = 0$, y que $\oint ds \mathbf{F}(\mathbf{r}(s)) \cdot \mathbf{r}'(s) = 0$.

Si la fuerza es función de la posición y la velocidad, lo anterior no se cumple.

Comentario 2.1.2. La fuerza debida a un campo magnético no genera trabajo

$$dW = \mathbf{F}(\mathbf{r}, \mathbf{t}) \cdot d\mathbf{x} = (q\dot{\mathbf{x}} \times \mathbf{B}(\mathbf{r}, \mathbf{t})) \cdot \dot{\mathbf{x}}dt = 0$$

Si el campo magnético varía en el tiempo según la ley de Faraday habrá una fuerza electromotriz no nula ($\mathcal{E} \neq 0$). Incluso si $\mathbf{B} = \mathbf{B}(\mathbf{r})$ la condición de campo de fuerza irrotacional no está definida.

El momento lineal tampoco es conservado, dado que la fuerza es en general perpendicular a la velocidad, por lo que una partícula con velocidad no nula cambiará de dirección de movimiento.

Llamando U a la *energía potencial dependiente de la velocidad* o *potencial generalizado*

$$\begin{aligned} d(T + U) &= F dx + \frac{\partial U}{\partial x} dx + \frac{\partial U}{\partial v} dv \\ &= F dx + \frac{\partial U}{\partial x} dx + d\left(\frac{\partial U}{\partial v} v\right) - v \frac{d}{dt} \frac{\partial U}{\partial v} dt \end{aligned}$$

⁶También llamado *momento conjugado* o *momento canónico*.

⁷Porque a cada punto de un subconjunto de \mathbb{R}^3 asigna un vector que representa a una propiedad física.

reemplazando según $v dt = dx$ y reordenando

$$d\left(T + U - \frac{\partial U}{\partial v}v\right) = \left(F + \frac{\partial U}{\partial x} - \frac{d}{dt} \frac{\partial U}{\partial v}\right) dx$$

Por lo que existirá una cantidad conservada si

$$F = -\frac{\partial U}{\partial x} + \frac{d}{dt} \frac{\partial U}{\partial v} \quad (2.3)$$

y se dice de $\mathbf{F}(\mathbf{r}, \mathbf{v})$ que es una *fuerza conservativa dependiente de la velocidad*.

2.1.5. Lagrangiano de una partícula en un campo magnético

En mecánica clásica, cuando las fuerzas son conservativas dependientes de la posición se cumple que $L = T - V$, y cuando son conservativas dependientes también de la velocidad se cumple $L = T - U$.

Comentario 2.1.3. Pueden probarse ambas igualdades reemplazándolas en la ecuación de movimiento (2.1). Naturalmente puede obtenerse la ecuación 2.3 introduciendo $L = T - U$ en la ec. 2.1.

El problema de encontrar L para una partícula sujeta a un campo eléctrico y magnético consiste en lo siguiente. Hallar la función U que mediante la ecuación 2.3 se obtenga la fuerza de Lorentz. Esa función es $U = z(\phi - \mathbf{v} \cdot \mathbf{A})$. La utilización del vector potencial en lugar del campo magnético no es arbitraria (no puede expresarse a U en términos de \mathbf{B} [1]) por lo que se pueden usar diferentes U dependiendo de la función *gauge* elegida. Esto tiene importantes consecuencias cuando los cálculos se realizan en forma aproximada.

2.1.6. Mecánica hamiltoniana

El lagrangiano no es una cantidad que siempre se conserve en los casos de interés práctico. Se define al *hamiltoniano* (H) del sistema

$$H(\mathbf{q}, \mathbf{p}, t) := \sum_i p \dot{q} - L$$

que se conserva cuando el sistema es invariante ante traslaciones temporales.

A partir de esta definición y de las ecuaciones de movimiento de Lagrange se puede derivar un formalismo alternativo para la descripción del sistema, el cual consiste en el cumplimiento de las *ecuaciones de movimiento de Hamilton*

$$\begin{aligned}\frac{\partial H}{\partial p_i} &= \dot{q}_i \\ \frac{\partial H}{\partial q_i} &= -\dot{p}_i \\ \frac{\partial H}{\partial t} &= -\frac{\partial L}{\partial t} \left(= \frac{dH}{dt} \right)\end{aligned}$$

Para un sistema de n coordenadas de posición se tienen $2n (+1)$ ecuaciones (de primer orden) de movimiento, a diferencia del formalismo de Lagrange en el que se tienen n ecuaciones de segundo orden.

2.1.7. Transformaciones de los potenciales

Dada una función escalar independiente de la velocidad $f(\mathbf{r}, t)$, las siguientes transformaciones *conjuntas* no afectan a los campos eléctrico y magnético

$$\begin{aligned}\mathbf{A}' &\longrightarrow \mathbf{A} + \nabla f(\mathbf{r}, t) \\ \phi' &\longrightarrow \phi + \frac{\partial f}{\partial t}\end{aligned}$$

ya que

$$\begin{aligned}\mathbf{E}(\phi', \mathbf{A}') &= -\nabla \phi' - \frac{\partial \mathbf{A}'}{\partial t} = -\nabla \left(\phi + \frac{\partial f}{\partial t} \right) - \frac{\partial (\mathbf{A} + \nabla f)}{\partial t} = \mathbf{E}(\phi, \mathbf{A}) \\ \mathbf{B}(\phi', \mathbf{A}') &= \nabla \times \mathbf{A}' = \nabla \times (\mathbf{A} + \nabla f) = \nabla \times \mathbf{A} + \nabla \times (\nabla f) = \nabla \times \mathbf{A} = \mathbf{B}(\phi, \mathbf{A})\end{aligned}$$

Estas transformaciones se llaman *transformaciones gauge* y la función $f(\mathbf{r}, t)$ *función gauge*.

2.1.8. Transformaciones en ecuaciones de movimiento

Las ecuaciones de Euler-Lagrange no se ven afectadas al transformar al lagrangiano de un sistema según

$$L' \longrightarrow L + \frac{df(\mathbf{r}, t)}{dt}$$

dado que la derivada temporal satisface idénticamente las ecuaciones de Euler-Lagrange, y estas son lineales en el lagrangiano⁸. Análogamente, dado que el nuevo lagrangiano sólo aporta un término constante ($f(\mathbf{r}(t_2), t_2) - f(\mathbf{r}(t_1), t_1)$) a la acción S , el valor estacionario de la acción $\delta S = 0$ seguirá correspondiendo a la misma trayectoria, ya que todas estas se ven afectadas aditivamente de la misma forma.

Como se indicó antes, el lagrangiano para una partícula en un campo electromagnético es $L = T - z(\phi - \mathbf{v} \cdot \mathbf{A})$. Por lo que

$$L' = T - z(\phi' - \mathbf{v} \cdot \mathbf{A}') = L + z\left(\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f\right)$$

El último término de esta ecuación es

$$\frac{df(\mathbf{r}, t)}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f$$

Por lo tanto las transformaciones *gauge* de los potenciales eléctrico y magnético no afectan a las ecuaciones de Euler-Lagrange. Las ecuaciones de Hamilton tampoco cambian su forma, pero el momento generalizado y el hamiltoniano se transforman según:

$$\begin{aligned}\mathbf{p}' &= \mathbf{p} + z\nabla f \\ H' &= H - z\frac{\partial f}{\partial t}\end{aligned}$$

2.2. Mecánica relativista

La descripción de los sistemas sometidos a campos magnéticos externos falla groseramente cuando se lleva a cabo aplicando las reglas de cuantización sobre el hamiltoniano clásico[2]. Esto se debe a que esta descripción no contempla una propiedad intrínseca de partículas subatómicas llamada *espín*. La misma es contemplada en forma natural en el tratamiento relativista. Por analogía con este, su introducción puede realizarse *ad hoc* para el caso clásico, y también como se describe con el Comentario 2.4.1. Se prefirió realizar, en forma sucinta, una descripción del tratamiento relativista por sobre aquel del Comentario 2.4.1. Esto otorga mayor coherencia con el formalismo seguido en este texto.

La teoría de la relatividad especial puede ser descrita con distintos conjuntos de axiomas. La forma más divulgada es aquella original de Einstein[3] que consta de dos postulados

⁸Representándolas con un mapa \mathcal{Q} , $\mathcal{Q}(L') = \mathcal{Q}(L) + \mathcal{Q}\left(\frac{df(\mathbf{r}, t)}{dt}\right) = \mathcal{Q}(L)$.

- 1 *Principio de Relatividad*: Las leyes que rigen la evolución de los sistemas físicos son independientes de si se refieren a uno u otro sistemas que estén en movimiento traslacional uniforme relativo entre ellos.
- 2 *Principio de constancia de la velocidad de la luz*: La luz en vacío se propaga a velocidad c (una constante fija, independiente de la dirección) en al menos un sistema de referencia inercial (el "sistema estacionario"), sea cual fuere el estado de movimiento de la fuente de luz.

Formas más modernas y elegantes han sido propuestas[4, 5, 6]

Comentario 2.2.1. En una forma más explícita el postulado 1 establece que las ecuaciones de movimiento, que describen al sistema en estudio, deben ser las mismas para ambos sistemas de referencia.

Es importante notar, que estos postulados tienen prioridad por sobre cualquier ley que describa algún tipo específico de interacción (ej. ecuaciones de Maxwell). De no hacerlo puede incurrirse en el error habitual de considerar innecesario al segundo postulado, dada la posibilidad de deducir la velocidad de la luz a partir de las ecuaciones de Maxwell.

Esto también está de acuerdo con el principio de relatividad de Galileo. Según el tratamiento de Palash[5] las transformaciones entre ambos sistemas de referencia se diferencian en ambas relatividades en una suposición extra. En el caso de Galileo es supuesto que el tiempo es el mismo en ambos sistemas de referencia, mientras que en la teoría de Einstein lo es la velocidad de la luz en vacío.

Puede demostrarse que de acuerdo con estos postulados las transformaciones de las coordenadas entre ambos sistemas deben ser las transformaciones de Lorentz y no las de Galileo.

Las ecuaciones de movimiento de Lagrange deben ser invariantes ante transformaciones de Lorentz. Para una partícula en un campo conservativo es válido

$$L = -mc^2\gamma^{-1} - V$$

donde γ representa al *factor de Lorentz*

$$\gamma := \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

y c a la velocidad de la luz en vacío.

Resolviendo las ecuaciones de movimiento para este lagrangiano se obtiene

$$\mathbf{F} = m \frac{d(\gamma \mathbf{v})}{dt} = -\nabla V(\mathbf{r})$$

Para una partícula en un campo electromagnético es válido el lagrangiano

$$L = -mc^2\gamma^{-1} - z(\phi - \mathbf{v} \cdot \mathbf{A})$$

dado que si es introducido en las ecuaciones de movimiento de Lagrange se obtiene la fuerza de Lorentz

$$\mathbf{F} = m \frac{d(\gamma \mathbf{v})}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

Si la expresión del momento conjugado (ver ecuación 2.2)

$$\mathbf{p} = m\gamma \mathbf{v} + z\mathbf{A}$$

es introducida en la función energía, se obtiene

$$h = \mathbf{p} \cdot \mathbf{v} - L = mc^2\gamma + z\phi$$

Si se utiliza al momento cinético $\boldsymbol{\pi} := m\gamma \mathbf{v} = \mathbf{p} - z\mathbf{A}$ la función energía puede ser reescrita en términos del momento conjugado, obteniéndose al hamiltoniano

$$\mathcal{H} = c\sqrt{\boldsymbol{\pi}^2 + m_e^2 c^2} + z\phi$$

2.2.1. Ecuación de Dirac

El hamiltoniano relativista de un electrón sometido a un campo electromagnético es

$$\mathcal{H} = c\sqrt{\boldsymbol{\pi}^2 + m_e^2 c^2} - e\phi \quad (2.4)$$

con $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}$. La ecuación de Schrödinger dependiente del tiempo particularizada a este hamiltoniano genera varios problemas. Al elevarla al cuadrado⁹ el operador energía $\mathcal{E} = i\hbar(\partial/\partial t)$ da origen a una derivada segunda respecto del tiempo. Esto tiene algunas implicancias que no eran aceptadas por los físicos. La relación entre el vector de estado $\Psi(x)$ y una amplitud de probabilidad sólo es posible si existe una densidad de probabilidad $\rho(x) = \Psi(x)^* \Psi(x)$ y una corriente de probabilidad $j(x) = \hbar/2mi(\Psi^*(\partial\Psi/\partial x) - \Psi(\partial\Psi^*/\partial x))$ (definición no válida si hay campos electromagnéticos¹⁰) que satisfagan la ecuación de continuidad

⁹Para quitar la raíz.

¹⁰En cuyo caso se le debe adicionar $-q/m\mathbf{A}|\psi|^2$.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$$

la cual garantizaría que no exista pérdida ni ganancia de probabilidad. Esto no se cumple para la ecuación en cuestión, incluso pueden surgir otros problemas como densidades de probabilidad negativa.

Dirac solucionó estos problemas transformando la ecuación 2.4 de modo que las derivadas respecto del espacio (asociada al operador momento) sean, al igual que la temporal, de primer orden. Dicha forma consiste en expresar como un cuadrado perfecto al argumento de la raíz cuadrada del hamiltoniano, de acuerdo con

$$\sum_{i=1}^3 \pi_i^2 + m_e^2 c^2 = \left(\sum_{i=1}^3 \alpha_i \pi_i + \alpha_0 m_e c \right)^2$$

Expandiendo al lado derecho y reagrupando

$$\sum_{i=1}^3 \pi_i^2 + m_e^2 c^2 = \sum_{i=1}^3 \alpha_i^2 \pi_i^2 + \alpha_0^2 m_e^2 c^2 + \sum_{i>j} (\alpha_i \alpha_j + \alpha_j \alpha_i) \pi_i \pi_j + \sum_{i=1}^3 (\alpha_i \alpha_0 + \alpha_0 \alpha_i) \pi_i m_e c \quad (2.5)$$

se consigue obtener la linealización buscada que, si se define a un operador β para la componente α_0 , puede expresarse como

$$\mathcal{H}_{\text{Dirac}} = c \boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \beta m_e c^2 - e\phi$$

Para satisfacer la igualdad 2.5 se requiere las relaciones

$$\alpha_i^2 = 1 \quad (2.6)$$

$$\alpha_i \alpha_j + \alpha_j \alpha_i = 0 \quad (2.7)$$

Estas últimas dos igualdades implican la relación de anticonmutación

$$\{\alpha_i, \alpha_j\} = 2\delta_{ij}$$

y también que los operadores $\{\alpha_k\}$ no pueden ser representados con escalares sino que necesitan matrices de dimensión mayor que 1 para ser representados, además que estas son de traza nula. Esto último se puede demostrar considerando $i \neq k$ con las siguientes ecuaciones

$$\begin{aligned}\text{tr } \alpha_i &= \text{tr}(\alpha_i 1) = \text{tr}(\alpha_i \alpha_k \alpha_k) = \text{tr}((0 - \alpha_k \alpha_i) \alpha_k) = -\text{tr}(\alpha_k (\alpha_i \alpha_k)) \\ &= -\text{tr}((\alpha_i \alpha_k) \alpha_k) = -\text{tr}(\alpha_i (\alpha_k \alpha_k)) = -\text{tr}(\alpha_i 1) = -\text{tr } \alpha_i\end{aligned}$$

Los autovalores de α_i son ± 1 , ya que si λ es un autovalor de α_i y f la correspondiente autofunción, entonces

$$f = 1f = \alpha_i^2 f = \lambda \alpha_i f = \lambda^2 f$$

Siendo así, dado que $\text{tr } \alpha_i = 0$ y que la traza de una matriz es igual a la suma de sus autovalores, la dimensión de α_i debe ser par. Por sencillez es conveniente evaluar la posibilidad de que sean de dimensión 2. Siendo que sus autovalores caben en una circunferencia de radio 1 en el plano complejo, las matrices son unitarias, que es una definición equivalente a decir que sus transpuestas conjugadas son iguales a sus inversas. Además, como las $\alpha_i^2 = 1$, α_i es su propia inversa, por lo que α_i es igual a su transpuesta conjugada (adjunta) y por lo tanto las α_i son hermíticas (y unitarias). Esta información permite encontrar la forma de las matrices.

De las infinitas representaciones posibles, se define a la *representación estándar* como aquella fijada por

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Las restantes α_i podrían obtenerse resolviendo la ecuación matricial $\{\alpha_i, \alpha_0\} = 0$, $i \neq 0$. Pero sólo puede encontrarse otras dos matrices que satisfagan esa relación, al sólo poder encontrar 3 de 4 matrices se asume que en realidad las matrices deben ser de 4×4 . Junto a la matriz anterior, las matrices de 2×2 encontradas con este procedimiento se llaman *matrices de Pauli* y son las siguientes

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Procediendo de esta forma ahora para matrices de 4×4 , es decir

$$0 = \alpha_i \beta + \beta \alpha_i = \begin{pmatrix} a & b \\ b^\dagger & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} a & b \\ b^\dagger & c \end{pmatrix} = \begin{pmatrix} 2a & 0 \\ 0 & -2c \end{pmatrix}$$

donde se utilizó forma general de una matriz hermítica para α_i se obtiene $a = c = 0_{2 \times 2}$.

Una solución posible es

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \alpha_{i>0} = \begin{pmatrix} 0 & \sigma \\ \sigma & 0 \end{pmatrix}$$

De esta manera el vector de estado tiene 4 componentes y la ecuación de Schrödinger independiente del tiempo puede expresarse mediante las ecuaciones

$$\begin{aligned} c((\pi_x - i\pi_y)\Psi_4 + \pi_z\Psi_3) &= (E - m_e c^2 + e\phi)\Psi_1 \\ c((\pi_x + i\pi_y)\Psi_3 - \pi_z\Psi_4) &= (E - m_e c^2 + e\phi)\Psi_2 \\ c((\pi_x - i\pi_y)\Psi_2 + \pi_z\Psi_1) &= (E + m_e c^2 + e\phi)\Psi_3 \\ c((\pi_x + i\pi_y)\Psi_1 - \pi_z\Psi_2) &= (E + m_e c^2 + e\phi)\Psi_4 \end{aligned}$$

Se desprecia el lado izquierdo de la igualdad por suponer que $m_e^2 c^4 \gg c^2 \pi^2$. Esto da origen a dos tipos de soluciones (para los autovalores), las electrónicas que son cercanas a $m_e c^2$, y las positrónicas que son cercanas a $-m_e c^2$. Ambos conjuntos de soluciones pueden desacoplarse, el procedimiento matemático es demasiado extenso como para incluirse aquí (ver referencia [7] para un tratamiento más completo), luego de aplicarlo, puede expresarse al hamiltoniano en series de c^{-2} resultando sus dos primeros miembros

$$\begin{aligned} \mathcal{H}^0 &:= \frac{1}{2m_e}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2 - e\phi \\ \mathcal{H}^1 &:= \frac{1}{8m_e^3 c^2}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^4 - \frac{ie\hbar}{8m_e^2 c^2} e [\boldsymbol{\sigma} \cdot \boldsymbol{\pi}, \boldsymbol{\sigma} \cdot \mathbf{E}] \end{aligned} \quad (2.8)$$

Aproximar hasta orden cero es suficiente para muchos casos prácticos.

2.3. Ecuación de Pauli

Se llama *ecuación de Pauli* a la ecuación de Schrödinger cuyo hamiltoniano es aquel de la ecuación 2.8

Por verificación directa puede demostrarse

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$$

donde ϵ_{ijk} es el *símbolo de Levi-Civita* definido por

$$\epsilon_{ijk} = \begin{cases} +1 & \text{si } (i, j, k) \text{ es } (1, 2, 3), (2, 3, 1) \text{ o } (3, 1, 2) \\ -1 & \text{si } (i, j, k) \text{ es } (3, 2, 1), (1, 3, 2) \text{ o } (2, 1, 3) \\ 0 & \text{si } i = j \text{ o } j = k \text{ o } k = i \end{cases}$$

Entonces

$$[\sigma_i, \sigma_j] + \{\sigma_i, \sigma_j\} = \delta_{ij} + i\epsilon_{ijk}\sigma_k$$

El operador $(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2$ puede desarrollarse según

$$\begin{aligned} (\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2 &= \left(\sum_i \sigma_i \pi_i \right) \left(\sum_i \sigma_i \pi_i \right) = \sum_{ij} \sigma_i \sigma_j \pi_i \pi_j = \sum_i \pi_i^2 + i \sum_{ijk} \epsilon_{ijk} \sigma_k \pi_i \pi_j \\ &= \pi^2 + i \boldsymbol{\sigma} \cdot \boldsymbol{\pi} \times \boldsymbol{\pi} \end{aligned}$$

Esta expresión puede simplificarse mucho, específicamente en cuanto a $\boldsymbol{\pi} \times \boldsymbol{\pi}$. Para ello conviene utilizar una función auxiliar f en el dominio del operador y utilizar la igualdad $\boldsymbol{\pi} = -i\hbar\nabla + e\mathbf{A}$

$$\begin{aligned} \boldsymbol{\pi} \times \boldsymbol{\pi} f &= (-i\hbar\nabla - e\mathbf{A}) \times (-i\hbar\nabla - e\mathbf{A}) f = (-i\hbar\nabla - e\mathbf{A}) \times (-i\hbar\nabla f - e\mathbf{A}f) = \\ &= -\nabla \times \nabla f + i\hbar e\mathbf{A} \times \nabla f + i\hbar e\nabla \times (\mathbf{A}f) + e^2\mathbf{A} \times \mathbf{A}f \end{aligned}$$

el primero y el último término del último miembro son cero dado que el rotor del gradiente es cero y también lo es el producto vectorial de un vector con él mismo, entonces

$$\begin{aligned} \boldsymbol{\pi} \times \boldsymbol{\pi} f &= i\hbar e\mathbf{A} \times \nabla f + i\hbar e\nabla \times (\mathbf{A}f) = i\hbar e\nabla \times \mathbf{A}(f) = i\hbar e\mathbf{B}f \\ \boldsymbol{\pi} \times \boldsymbol{\pi} &= i\hbar e\mathbf{B} \end{aligned}$$

Entonces

$$\frac{(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2}{2m_e} = \frac{\pi^2}{2m_e} + \frac{e\hbar}{2m_e} \mathbf{B} \cdot \boldsymbol{\sigma}$$

2.4. Espín

El operador de *espín* (\mathbf{s})

$$\mathbf{s} := \frac{\hbar}{2} \boldsymbol{\sigma}$$

tiene componentes s_k que satisfacen

$$[s_i, s_j] = i\hbar \varepsilon_{ijk} s_k$$

Nótese que la ecuación anterior corresponde a las relaciones de conmutación del momento angular. Como los autovalores de σ_i son ± 1 , los de s_i son $\pm \hbar/2$ que representan las dos componentes de un estado con número cuántico de momento angular $1/2$, y momento angular total ($s = \sqrt{s_x^2 + s_y^2 + s_z^2}$) $\sqrt{3/4}\hbar$.

Al establecer la representación matricial se lo hizo de modo que los vectores base (cuyas representaciones matriciales son $[1, 0]^T$ y $[0, 1]^T$) sean autovectores de s^2 y s_z .

$$\begin{aligned} s^2 \alpha &= \frac{3\hbar^2}{4} \alpha & s_z \alpha &= \frac{\hbar}{2} \alpha \\ s^2 \beta &= \frac{3\hbar^2}{4} \beta & s_z \beta &= -\frac{\hbar}{2} \beta \end{aligned}$$

Nótese que el estado de espín (autovector de s_z) tiene autovalor $m_s \hbar$, m_s es llamado *número cuántico de espín*, vale $1/2$ o $-1/2$ para autoestados de espín α y β respectivamente.

La parte espacial del vector de estado pertenece a un espacio de Hilbert separable de dimensión infinita. La parte de espín es representada por un vector en el espacio de Hilbert de espín de dimensión finita, para el caso de un electrón es de dimensión dos, y cuya base ortogonal está conformada por α y β . Para adecuar esto al tratamiento usado habitualmente en química, dado en términos de funciones, se representa a los elementos de este espacio de espín con funciones, y se utilizan como base a las funciones de espín definidas por

$$\begin{aligned} \alpha(1/2) &= 1 & \beta(1/2) &= 0 \\ \alpha(-1/2) &= 0 & \beta(-1/2) &= 1 \end{aligned}$$

las cuales son ortogonales, de acuerdo con las propiedades de los vectores que representan.

De esta forma con una combinación lineal de estas dos funciones se puede representar a todos los estados de espín de un electrón, incluyendo a los de dirección x e y . Para una dirección arbitraria determinada por un vector unitario \mathbf{n} , el operador de espín toma la forma $\mathbf{n} \cdot \mathbf{s}$.

Formalmente el espacio de soluciones se construye a partir del producto tensorial del espacio de Hilbert asociado a la parte espacial con el espacio de espín. En la práctica

se define a este producto tensorial de modo que las funciones de onda correspondan al producto de las funciones espacial y de espín.

Comentario 2.4.1. De lo anterior no se puede deducir que el espín sea un efecto relativista. De hecho, no lo es, aunque existe en la bibliografía cierta confusión al respecto dado que el espín surge naturalmente de la ecuación de Dirac. El espín también puede surgir de una manera más general, pero se escogió exponer la descripción anterior por ser la más sencilla entre las que se adecúan a los fines de este capítulo.

Un tratamiento más general puede encontrarse por ejemplo en la referencia [8], capítulos 3 y 7, cuya línea argumentativa puede resumirse del siguiente modo: variables físicas, entre ellas el momento angular, están relacionadas con transformaciones de simetría. Estas transformaciones pueden asociarse a dos sistemas de referencia con movimiento uniforme entre ellos. Aunque rigurosamente debieran ser transformaciones de Lorentz, a velocidades mucho menores que las de la luz puede usarse transformaciones de Galileo. Estas transformaciones son elementos de un grupo de Galileo. Este grupo tiene diez parámetros, tres ángulos de rotación, tres desplazamientos espaciales, tres componentes de velocidad y un parámetro de desplazamiento temporal. A estas operaciones se les puede asociar operadores unitarios, nótese que su aplicación a la función de onda no debe variar los valores de los observables físicos. Estos operadores unitarios pueden ser expresados como exponenciales cuyos argumentos contienen el producto de un parámetro y un generador hermitico. Estos últimos son luego relacionados con las variables dinámicas. Se encuentra entonces que el operador de momento angular J puede relacionarse con el de posición y momento mediante $\mathcal{J} = \mathcal{Q} \times \mathcal{P}$. Se plantea la posibilidad de existencia de un grado de libertad interno (el espín), independiente por tanto del centro de masa y consecuentemente de \mathcal{Q} y \mathcal{P} , por lo que su operador conmuta con estos. En cuyo caso debe cumplirse $\mathcal{J} = \mathcal{Q} \times \mathcal{P} + \mathcal{S}$. Sus propiedades matemáticas son determinadas naturalmente en el marco de este tratamiento.

Debe tenerse en cuenta que esencialmente se obtienen las propiedades que debiera tener el operador momento angular según propiedades de simetrías de rotación, y el espín surge cuando se generaliza esto considerando la posibilidad de un grado de rotación independiente del centro de masa.

2.5. Ideas principales sobre el capítulo

El objetivo específico perseguido con la escritura de este capítulo es describir una forma de construcción de un hamiltoniano *clásico* genérico que incluya las interacciones magnéticas. Como referencia se utilizó principalmente los textos de las referencias 1, 9, 10 y 11.

Para ello se resumió una versión moderna de los postulados de la mecánica newtoniana, contemplando fuerzas conservativas dependientes de la velocidad. Esto último es esencial en el tratamiento de interacciones magnéticas.

Se describió al formalismo de la mecánica lagrangiana y su aplicación a sistemas que, en el caso de la mecánica newtoniana, se describen con fuerzas conservativas dependientes de la velocidad; particularmente al caso de una partícula puntual cargada en un campo

magnético. Para poder hacer esto se describió al potencial eléctrico y al potencial vector magnético. Lo anterior se expuso dado que, como se mostró, el hamiltoniano de un sistema se obtiene a partir de la aplicación de una transformada de Legendre sobre el lagrangiano.

Se mostró la arbitrariedad de las definiciones de los potenciales, en el sentido de la invarianza de las ecuaciones de movimiento ante las transformaciones *gauge*. Este hecho tiene interés en la realización de cálculos aproximados de propiedades magnéticas.

Dado que a partir del hamiltoniano clásico y las reglas usuales de cuantización no se obtiene un hamiltoniano cuántico que describa correctamente partículas inmersas en campos magnéticos, se describió un tratamiento relativista en el que naturalmente surge el momento angular de espín. Se escogió esta modalidad por ser la más sencilla.

Finalmente y sin pretensiones históricas se describió a la ecuación de Pauli como aproximación de orden cero de la ecuación de Dirac, realizando manipulaciones algebraicas para obtener una expresión en términos del campo magnético.

Capítulo 3

Propiedades magnéticas

3.1. Hamiltoniano de espín efectivo

En la espectroscopía RMN se registra la absorción de radiación electromagnética por parte de moléculas inmersas en un campo magnético uniforme. Dicha absorción puede ser interpretada a partir de transiciones entre autoestados de espín.

El análisis no se realiza utilizando al hamiltoniano de espín obtenido teóricamente, que fue descrito en el capítulo anterior, dado que esto es actualmente irrealizable. Con fines prácticos e interpretativos se utiliza un hamiltoniano más sencillo llamado *hamiltoniano efectivo de espín*. Este hamiltoniano no incluye ninguna variable electrónica. En él los núcleos son descritos por sus correspondientes espines y sus momentos magnéticos asociados. Las interacciones quedan descritas en forma menos precisa en términos de constantes de apantallamiento y las constantes de acoplamiento espín-espín indirecto. En caso de no existir libre rotación, también es necesario incluir en este hamiltoniano a constantes de acoplamiento espín-espín directo.

Comentario 3.1.1. Un hamiltoniano efectivo es uno que actúa en un subespacio del espacio en el que actúa el hamiltoniano verdadero. En particular, el hamiltoniano de espín para RMN no actúa sobre coordenadas espaciales de las partículas del sistema. Por hamiltoniano verdadero debe entenderse aquí a alguno que incluya todas las interacciones conocidas, o al menos todas aquellas con un efecto notable para el tema en cuestión. Nótese que su definición puede ser dependiente del contexto, por ejemplo, incluir o no efectos relativistas.

Una condición requerida en la construcción de un hamiltoniano efectivo es la posibilidad de separación del hamiltoniano verdadero, dividiendo al problema en un espacio interno (de especial interés) y uno externo que describa las interacciones que no se pretenden describir, cuyos aportes a la energía deben ser notablemente diferentes.

Si bien las coordenadas electrónicas no son incluidas *explícitamente* en el hamiltoniano de espín efectivo, sus efectos sí son contemplados mediante los parámetros incluidos en él. Estos son ajustados (numéricamente) para concordar con las energías observadas experimentalmente. Se puede realizar lo anterior con relativamente poco costo computacional, debido a su reducido número de parámetros.

Es muy importante comprender que los hamiltonianos efectivos tienen la ventaja de simplificar el problema en términos interpretativos, permitiendo realizar representaciones mentales que serían imposibles de otro modo. Por ejemplo, las constantes de apantallamiento que son interpretadas como cuantificaciones de la reducción de la intensidad del campo magnético externo neto sufrido por los núcleos debida a los electrones.

3.1.1. Descripción fenomenológica

La contribución de un núcleo, a las señales obtenidas en RMN, es debida a su momento dipolar magnético (μ). Este se relaciona con su momento angular de espín (I) según

$$\mu = \hbar\gamma I$$

donde γ representa a la relación giromagnética del núcleo. Si dicho núcleo se encuentra aislado y sometido a un campo magnético uniforme \mathbf{B} , se generará una diferencia de energía potencial $\Delta U = -\mu \cdot \mathbf{B}$.

Comentario 3.1.2. Esta inclusión puede no resultar trivial. En primer término, desde un punto de vista clásico, la fuerza sobre una partícula cargada asociada al campo magnético no realiza trabajo (ver Comentario 2.1.2). Clásicamente un dipolo magnético está muy bien representado por un anillo por el cual circula una corriente. Tendrá un momento dipolar μ que apuntará en dirección perpendicular a las direcciones de la corriente. Es un resultado bien conocido que en presencia de un campo magnético uniforme habrá una fuerza que tienda a alinear al dipolo con el campo, dada por $\mathbf{F} = \nabla(\mu \cdot \mathbf{B})$ [9]. Esta es asociada inmediatamente a un cambio de energía potencial $\Delta U = -\mu \cdot \mathbf{B}$.

Un hecho que no suele ser discutido correctamente en la bibliografía es que, dado que la corriente se debe a partículas portadoras de carga, y que la fuerza asociada al campo magnético no realiza trabajo, el cambio en la energía potencial no puede ser debido al campo magnético. Dicho cambio puede justificarse considerando la ley de Faraday¹,

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt} \int \mathbf{B} \cdot d\mathbf{a}$$

Al rotar la espira cambia el flujo del campo magnético a través de ella. Según esta ley, esto implica la existencia una fuerza electromotriz. Por lo tanto, un cambio neto de energía potencial cuando las partículas portadoras de carga completan un ciclo al anillo. Este cambio en la energía se ve compensado por el cambio en la energía cinética de las partículas, conservándose así la energía mecánica total.

¹Esta es una de las ecuaciones de Maxwell. Se la expresó en término de integrales. Esta expresión puede obtenerse aplicando el teorema de Stokes sobre la forma dada en la sección de campos electromagnéticos.

En el caso de átomos y moléculas, el momento magnético tiene dos orígenes: el movimiento orbital y el espín. El primero puede tratarse análogamente al caso clásico recién comentado[10]. Para este, en el marco formal de la mecánica cuántica, puede demostrarse que: el incremento del valor esperado de la energía cinética traslacional de un átomo, conlleva un decrecimiento igual en magnitud, en el valor esperado de la energía potencial.

El momento magnético de espín surge² como un momento magnético intrínseco de las partículas. En apariencia la descripción previa no es aplicable a este momento magnético. Lo anterior parece una clara contradicción con el tratamiento clásico en el que el campo magnético no realiza trabajo. Esto podría reconciliarse si la partícula tuviese una energía cinética rotacional intrínseca. Ha sido demostrado que esto último es compatible con la ecuación de Dirac[11]. Esta idea ha sido extendida considerando que la energía en reposo (mc^2) es debida a la energía cinética de espín[12]

Existen también otras contribuciones a la energía del sistema. La interacción *directa* entre los momentos magnéticos de un par de núcleos también aporta a su energía. Esta puede interpretarse clásicamente como la energía potencial cuyo gradiente cambiado de signo representa la fuerza que tiende a alinear a ambos dipolos.

El potencial vector magnético puede expandirse en una serie multipolar, en forma análoga al campo eléctrico. El segundo término representa a la contribución dipolar. Conservando sólo a este término puede obtenerse una aproximación del potencial vector magnético en términos del momento dipolar magnético

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi r^2} \boldsymbol{\mu} \times \mathbf{r}$$

donde μ_0 es la permeabilidad magnética del vacío. Con manipulaciones algebraicas puede obtenerse la correspondiente expresión para el campo magnético de un dipolo situado en el origen de coordenadas[9]

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi r^3} [3(\boldsymbol{\mu} \cdot \mathbf{r})\mathbf{r} - \boldsymbol{\mu}]$$

Sea el campo anterior aquel generado por el núcleo 2. Para obtener la energía del acoplamiento debe multiplicarse por el inverso aditivo del momento magnético del núcleo 1. Realizando este producto, la expresión en términos de los momentos angulares de espín nucleares resulta

$$H_{\text{dip-dip}} = -\frac{\mu_0 \gamma_1 \gamma_2 \hbar^2}{4\pi r^3} [3(\mathbf{I}_1 \cdot \mathbf{r})(\mathbf{I}_2 \cdot \mathbf{r}) - \mathbf{I}_1 \cdot \mathbf{I}_1]$$

Utilizando coordenadas cartesianas, puede expresarse matricialmente como

²Por ejemplo, con el tratamiento de de Dirac.

$$H_{\text{dip-dip}} = \frac{\mu_0 \gamma_1 \gamma_2 \hbar^2}{4\pi r^3} \begin{pmatrix} I_{1,x} & I_{1,y} & I_{1,z} \end{pmatrix} \cdot \begin{pmatrix} 1 - 3\frac{x^2}{r^2} & -3\frac{xy}{r^2} & -3\frac{xz}{r^2} \\ -3\frac{xy}{r^2} & 1 - 3\frac{y^2}{r^2} & -3\frac{yz}{r^2} \\ -3\frac{xz}{r^2} & -3\frac{yz}{r^2} & 1 - 3\frac{z^2}{r^2} \end{pmatrix} \cdot \begin{pmatrix} I_{2,x} \\ I_{2,y} \\ I_{2,z} \end{pmatrix}$$

Nótese que la traza de esta matriz cuadrada (**D**) es nula, lo cual tiene importantes consecuencias. En experimentos de RMN en fases líquida o gaseosa, las moléculas no tienen una orientación de preferencia en el espacio. Por lo tanto, aquello medido corresponde al promedio de las distintas orientaciones en el espacio. Sea Ω una magnitud dependiente de su orientación espacial. Dado un sistema de ejes cartesianos, Ω puede expresarse para algún sistema rotado respecto del primero, a partir de los ángulos de Euler, mediante una función de estos: $\Omega(\alpha, \beta, \gamma)$. Si las orientaciones son equi-probables, entonces el promedio rotacional se obtiene mediante la siguiente operación lineal

$$\langle \Omega \rangle = \frac{1}{8\pi^2} \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \Omega(\alpha, \beta, \gamma) \sin(\alpha) d\alpha d\beta d\gamma$$

Que la traza de la matriz **D** sea nula implica que el promedio rotacional de su aporte a la energía es cero[13]. Al no haber encontrado una demostración de dicha afirmación, la siguiente demostración es propuesta³

Demostración 3.1.3. Es de interés que el promediado rotacional de la energía debido a la matriz **D** tiene la forma $cte. \times \langle \mathbf{I}_1 \cdot \mathbf{D} \cdot \mathbf{I}_2 \rangle$. Los \mathbf{I}_i pueden expandirse en una base ortonormal cuyos vectores coincidan con los ejes cartesianos

$$\mathbf{I}_j = i_{j,x} \mathbf{e}_x + i_{j,y} \mathbf{e}_y + i_{j,z} \mathbf{e}_z$$

Entonces, por la linealidad recién mencionada

$$\langle H_{\text{dip-dip}} \rangle = cte. \sum_{k,l} i_{1,k} i_{2,l} \langle \mathbf{e}_k \cdot \mathbf{D} \cdot \mathbf{e}_l \rangle \quad (3.1)$$

Cualquier rotación puede ser entendida como el efecto de un operador unitario $\Gamma^{(4)}$ al ser aplicado sobre **D**, en la forma

$$\mathbf{D}' = \mathbf{\Gamma}^T \mathbf{D} \mathbf{\Gamma}$$

Entonces, cada expresión del tipo $\langle \mathbf{e}_k \cdot \mathbf{D} \cdot \mathbf{e}_l \rangle$ en 3.1 luego de una rotación cumplirá

³El autor considera improbable que la demostración sugerida presente alguna originalidad.

⁴Con determinante igual a 1.

$$\langle \mathbf{e}_k \cdot \mathbf{D}' \cdot \mathbf{e}_l \rangle = \langle \mathbf{e}_k \cdot \mathbf{\Gamma}^T \mathbf{D} \mathbf{\Gamma} \cdot \mathbf{e}_l \rangle = \sum_{i,j} D_{i,j} \langle \Gamma_{k,i} \Gamma_{l,j} \rangle$$

Los $\Gamma_{k,i}$ pueden ser representados por los elementos de la matriz de cosenos directores en términos de los ángulos de Euler. El promediado rotacional para cada par de componentes $\langle \Gamma_{k,i} \Gamma_{l,j} \rangle$ puede realizarse con la integral anterior, reemplazando $\Omega(\alpha, \beta, \gamma)$ con $\Gamma_{k,i} \Gamma_{l,j}$. Lo recién descrito cumple que

$$\langle \Gamma_{k,i} \Gamma_{l,j} \rangle = \frac{1}{3} \delta_{kl} \delta_{ij}$$

Entonces

$$\langle \mathbf{e}_k \cdot \mathbf{D}' \cdot \mathbf{e}_l \rangle = \sum_{i,j} D_{i,j} \frac{1}{3} \delta_{kl} \delta_{ij}$$

Si $k \neq l$, entonces $\langle \mathbf{e}_k \cdot \mathbf{D}' \cdot \mathbf{e}_l \rangle = 0$. Si $k = l$, entonces $\langle \mathbf{e}_k \cdot \mathbf{D}' \cdot \mathbf{e}_l \rangle = \frac{\text{tr} \mathbf{D}}{3}$. Pero como se mostró, según la definición de \mathbf{D} , su traza es nula. Por lo tanto

$$\langle H_{\text{dip-dip}} \rangle = 0$$

□

Si bien es nulo el aporte a la energía del promediado rotacional de la interacción directa entre los momentos magnéticos de espín de dos núcleos, existe una interacción indirecta mediada por los electrones. El momento magnético de un primer núcleo tiene interacción con los momentos magnéticos orbital y de espín de los electrones, y estos a su vez con el momento magnético de un segundo núcleo. Se considera que estas interacciones tienen la misma forma que las del acoplamiento directo, y son descritas por el *tensor de acoplamiento indirecto espín-espín* (\mathbf{J}). Es decir, para cada par de núcleos hay un aporte a la energía proporcional a

$$\mathbf{I}_1 \cdot \mathbf{J}_{1,2} \cdot \mathbf{I}_2$$

Es importante notar que \mathbf{J} depende directamente de las relaciones giromagnéticas de los núcleos. Normalmente resulta conveniente trabajar con el *tensor de acoplamiento indirecto espín-espín reducido* (\mathbf{K}), que es definido por la siguiente igualdad

$$\mathbf{J}_{1,2} = h \frac{\gamma_1 \gamma_2}{4\pi^2} \mathbf{K}_{1,2}$$

Las componentes del tensor de acoplamiento indirecto espín-espín resultan independientes de las relaciones giromagnéticas de los núcleos. Por lo tanto, reflejan en una forma más directa el efecto de los electrones.

En el caso de \mathbf{K} , no hay razón para suponer que se anulará en el promediado rotacional, ya que al desconocer su forma no se puede afirmar que su traza sea nula. De hecho, resulta no serlo. Al realizar el promediado rotacional se obtiene, por similitud con \mathbf{D} , que su aporte a la energía puede reducirse a

$$\frac{1}{2} \frac{\text{tr } \mathbf{K}}{3} \boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2$$

Esto permite simplificar en mucho el análisis de los experimentos, ya que por cada par de núcleos sólo debe ajustarse un único valor en vez de las 3×3 componentes de la representación matricial del correspondiente tensor.

En la posición de un dado núcleo el campo magnético no corresponde en su intensidad al externo, debido a los campos magnéticos generados por el resto de las partículas. Se dice entonces que el campo magnético externo está apantallado en la posición nuclear. A dicho apantallamiento se asigna *tensor de apantallamiento* σ . Y la expresión aceptada para la energía de la interacción del momento magnético nuclear, de un dado núcleo, con el campo magnético externo es

$$H = -\boldsymbol{\mu}^T \cdot (\mathbf{1}_{3 \times 3} - \sigma) \cdot \mathbf{B}$$

Esta componente de la energía también debe ser promediada rotacionalmente. Para esto se puede utilizar una simplificación de la demostración propuesta para el caso de \mathbf{D} .

Demostración 3.1.4. Considérese el campo magnético uniforme alineado con la dirección z según los ejes del laboratorio. Esto permite considerar sólo a la componente z tanto de \vec{B} como de $\boldsymbol{\mu}$. Aplicando la demostración propuesta y la igualdad $\text{tr}(A + B) = \text{tr } A + \text{tr } B$, se obtiene

$$\langle H \rangle = -\mu_z(1 - \sigma)B_z$$

donde σ es un tercio de la traza de σ . □

Entonces, para una molécula en libre rotación sometida a un campo magnético con componente no nula únicamente en el eje z , el hamiltoniano de espín efectivo es

$$\mathcal{H} = -\sum_k \gamma_k \hbar (1 - \sigma_k) B_z I_{k,z} + \sum_{k>l} \gamma_k \gamma_l \hbar^2 K_{kl} \mathbf{I}_k \cdot \mathbf{I}_l$$

3.2. Parámetros del hamiltoniano efectivo de espín y derivadas de la energía

El hamiltoniano de espín efectivo utilizado para la interpretación de espectros de RMN es diferente de aquel derivado a partir de primeros principios. El primero, debido a su forma, puede ser vinculado con otros hamiltonianos. Nótese que al expandir la energía en series de Taylor alrededor de $(\mathbf{B}, \boldsymbol{\mu}) = (\mathbf{0}, \mathbf{0})$

$$E(\mathbf{B}, \boldsymbol{\mu}) = E_0 + \frac{1}{2} \mathbf{B}^T E^{20} \mathbf{B} + \sum_k \mathbf{B}^T E_k^{11} \boldsymbol{\mu}_k + \frac{1}{2} \sum_{k \neq j} \boldsymbol{\mu}_k^T E_{kj}^{02} \boldsymbol{\mu}_j + \dots$$

donde se denotó con $\boldsymbol{\mu}$ al conjunto de los momentos magnéticos nucleares. Por comparación con el hamiltoniano de espín efectivo puede deducirse que

$$\begin{aligned} \sigma_k &= E_k^{11} + 1 \\ K_{kl} &= E_{kl}^{02} - D_{kl} \end{aligned}$$

Para que estas expresiones sean de utilidad, no deben generarse errores importantes por el truncado de la serie. En unidades atómicas, la magnitud de los $\boldsymbol{\mu}_k$ es del orden de 10^{-4} , al igual que la de las constantes de acoplamiento directo. En las experiencias de resonancia magnética, la magnitud de \mathbf{B} suele ser del orden de $4,3 \times 10^{-5}$ (10 T). Las transiciones resultan entonces del orden de 10^{-8} (10^{-13} para los multipletes). Estos cambios energéticos son muy pequeños en comparación con las energías asociadas a enlaces químicos, cuyo orden es 10^{-1} . La pequeñez de los $\boldsymbol{\mu}_k$ y \mathbf{B} permite despreciar los términos de orden superior en el desarrollo de Taylor.

3.3. Parámetros de RMN para funciones de onda exactas

La energía de un sistema que varía ante una perturbación puede ser expresada como una función de la misma. El cambio en la energía generado por la perturbación puede expresarse como mediante una serie de Taylor. Los coeficientes correspondientes a las derivadas de la energía en respecto de la variable asociada a la perturbación, son llamados *propiedades moleculares*. Se las clasifican en primer orden, segundo orden etc., en coincidencia con el orden de las derivadas a las que corresponden.

Con teoría de perturbaciones para estados estacionarios no degenerados, las propiedades moleculares de primer y segundo orden tienen las siguientes expresiones

$$\begin{aligned}\frac{dE(\mathbf{x})}{dx_i} &= \langle 0 | \frac{d\mathcal{H}}{dx_i} | 0 \rangle \\ \frac{d^2E(\mathbf{x})}{dx_i dx_j} &= \langle 0 | \frac{d^2\mathcal{H}}{dx_i dx_j} | 0 \rangle - 2 \sum_{n \neq 0} \frac{\langle 0 | \frac{d\mathcal{H}}{dx_i} | n \rangle \langle n | \frac{d\mathcal{H}}{dx_j} | 0 \rangle}{E_n - E_0}\end{aligned}$$

donde las derivadas son evaluadas en $\mathbf{x} = \mathbf{0}$. La expresión para las propiedades moleculares de primer orden corresponde al teorema de Hellmann-Feynman. La contribución del valor esperado a la propiedad de segundo orden es llamada parte *diamagnética*. A la contribución de la suma de estados se le llama parte *paramagnética*.

Es importante destacar que estas expresiones requieren el conocimiento de los estados excitados, lo cual es muy poco práctico desde el punto de vista del cálculo.

3.3.1. Hamiltoniano molecular y forma explícita del campo magnético

El Hamiltoniano electrónico no relativista de un sistema electrónico molecular sometido a un campo magnético externo es

$$\begin{aligned}\mathcal{H} &= \frac{1}{2} \sum_i \pi_i^2 - \sum_i^{N_e} \boldsymbol{\mu}_i \cdot \mathbf{B}^{tot}(\mathbf{r}_i) - \sum_{iK} \frac{Z_K}{r_{iK}} + \frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}} \\ &+ \frac{1}{2} \sum_{KL} \frac{Z_K Z_L}{R_{KL}} - \sum_K \boldsymbol{\mu}_K \cdot \mathbf{B}^{tot}(\mathbf{R}_K) + \sum_{K>L} \boldsymbol{\mu}_K^T \mathbf{D}_{KL} \boldsymbol{\mu}_L\end{aligned}\quad (3.2)$$

con $\pi_j = -i\nabla_j + \mathbf{A}^{tot}(\mathbf{r}_j)$. $\mathbf{A}^{tot}(\mathbf{r}_j)$ es el potencial vector magnético en la posición \mathbf{r}_j debido al campo magnético externo ($\mathbf{A}_O(\mathbf{r}_j)$) y al generado por los núcleos, que corresponde a la sumatoria sobre el asociado a cada núcleo K ($\mathbf{A}_K(\mathbf{r}_j)$).

Para poder realizar cálculos es necesario conocer las expresiones explícitas de las variables que aparecen en él. Para esto es necesario elegir un *gauge*. El más usado para este tipo de cálculos es el de Coulomb, que consiste en imponer $\nabla \cdot \mathbf{A} = 0$. Con él, el potencial vector asociado a un campo magnético uniforme toma la forma

$$\mathbf{A}_O(\mathbf{r}_i) = \frac{1}{2} \mathbf{B} \times \mathbf{r}_{iO}$$

Para este *gauge*, el potencial vector magnético generado por un núcleo es

$$\mathbf{A}_K(\mathbf{r}_i) = \alpha^2 \frac{\boldsymbol{\mu}_K \times \mathbf{r}_{iO}}{r_{iK}^3}$$

y puede demostrarse entonces que el campo magnético correspondiente es

$$\mathbf{B}_K(\mathbf{r}_i) = -\alpha^2 \frac{r_{iK}^2 \mathbf{1} - 3\mathbf{r}_{iK}\mathbf{r}_{iK}^T}{r_{iK}^5} \boldsymbol{\mu}_K + \frac{8\pi\alpha^2}{3} \delta(\mathbf{r}_{iK}) \boldsymbol{\mu}_K$$

En las expresiones anteriores, los segundos subíndices O y K describen el origen de coordenadas. K representa que el origen de coordenadas corresponde con la posición del núcleo en cuestión. O indica que no existe una obvia preferencia por la elección del origen de coordenadas.

3.3.2. Expresiones explícitas de propiedades moleculares

Las expresiones de las propiedades moleculares suelen ser separadas en suma de términos, de acuerdo con las interacciones que con ellos se describe. Las propiedades de primer orden, correspondientes a las perturbaciones dadas por \mathbf{B} son

$$\begin{aligned} \frac{d\mathcal{H}}{d\mathbf{B}} &= \mathbf{h}_B^{orb} + \mathbf{h}_B^{spn} \\ \mathbf{h}_B^{orb} &:= \frac{1}{2} \sum \mathbf{I}_{jO} \\ \mathbf{h}_B^{spn} &:= \sum \mathbf{s}_j \end{aligned}$$

En estas expresiones $\mathbf{I}_{jO} = -i\mathbf{r}_{jO} \times \nabla_j$ es el operador de momento angular orbital para el electrón j , y \mathbf{s}_j su operador de espín. El acoplamiento de primer orden entre el campo magnético y los electrones es llamado *interacción de Zeeman*.

Las propiedades de primer orden, correspondientes a las perturbaciones dadas por μ son

$$\begin{aligned}
\frac{d\mathcal{H}}{d\boldsymbol{\mu}_K} &= \mathbf{h}_K^{pso} + \mathbf{h}_K^{sd} + \mathbf{h}_K^{fc} \\
\mathbf{h}_K^{pso} &:= \alpha^2 \sum_i \frac{\mathbf{I}_{iK}}{r_{iK}^3} \\
\mathbf{h}_K^{sd} &:= \alpha^2 \sum_i \frac{r_{iK}^2 \boldsymbol{\mu}_i - 3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{iK}) \mathbf{r}_{iK}}{r_{iK}^5} \\
\mathbf{h}_K^{fc} &:= \frac{8\pi\alpha^2}{3} \delta(\mathbf{r}_{iK}) \boldsymbol{\mu}_i
\end{aligned}$$

El operador \mathbf{h}_K^{pso} corresponde al acoplamiento entre los momentos magnéticos nucleares y momento angular orbital de los electrones, y es llamado *operador paramagnético espín-órbita* u *operador hiperfino*. Los operadores \mathbf{h}_K^{sd} y \mathbf{h}_K^{fc} son llamados *operador espín-dipolar* y *operador del contacto de Fermi*, respectivamente. El primero, al ser multiplicado por el momento magnético nuclear, representa la interacción entre dos dipolos clásicos. El operador de contacto de Fermi representa la interacción del momento dipolar del electrón con la fuente de campo magnético nuclear cuando el electrón se encuentra en la posición del núcleo.

Las propiedades moleculares de segundo orden

$$\begin{aligned}
\frac{d^2\mathcal{H}}{d\mathbf{B}d\boldsymbol{\mu}_K} &= -\mathbf{1} + \mathbf{h}_{BK}^{diag} \\
\frac{d^2\mathcal{H}}{d\boldsymbol{\mu}_K d\boldsymbol{\mu}_L} &= \mathbf{D}_{KL} + \mathbf{h}_{KL}^{dso}
\end{aligned}$$

son expresadas en términos de los operadores

$$\begin{aligned}
\mathbf{h}_{BK}^{diag} &:= \frac{\alpha^2}{2} \sum_i \frac{(\mathbf{r}_{iO} \cdot \mathbf{r}_{iK}) \mathbf{1} - \mathbf{r}_{iK} \cdot \mathbf{r}_{iO}^T}{r_{iK}^3} \\
\mathbf{h}_{KL}^{dso} &:= \frac{\alpha^4}{2} \sum_i \frac{(\mathbf{r}_{iK} \cdot \mathbf{r}_{iL}) \mathbf{1} - \mathbf{r}_{iK} \cdot \mathbf{r}_{iL}^T}{r_{iL}^3 r_{iK}^3}
\end{aligned}$$

Ambas propiedades son expresadas como suma de dos términos. Uno asociado con los núcleos; en el primer caso por diferenciación de la interacción de Zeeman y en el segundo caso asociado a la interacción entre dos dipolos clásicos. Y otro término asociado a interacciones electrónicas (\mathbf{h}_{BK}^{diag} y \mathbf{h}_{KL}^{dso} respectivamente). La existencia en estos operadores de los factores α^2 o α^4 implica que sólo son pequeños respecto a las contribuciones puramente nucleares. Sendos nombres son *operador diamagnético* y *operador diamagnético espín-órbita*.

3.4. Cálculo de las propiedades magnéticas

Las propiedades magnéticas utilizadas en esta tesis para estudios QSAR/QSPR fueron expuestas en la sección anterior. En esta sección se describen someramente ideas generales acerca de cómo se realizan estos cálculos.

3.4.1. Principios generales para el cálculo en funciones de onda aproximadas

Las expresiones obtenidas para las propiedades magnéticas no son útiles en la práctica. Esto es así porque requieren la expresión explícita de los estados excitados. Una forma de evitar esto es mediante la teoría de perturbaciones variacional. Esta puede ser formulada tanto para métodos variacionales como para métodos no variacionales. Se considerará a continuación el caso variacional.

Para algún método dado, puede considerarse a la energía como una funcional del campo magnético, los momentos magnéticos y algún conjunto de parámetros λ , respecto de los cuales, la función de onda es variacional. Se impone la condición variacional

$$\left. \frac{\partial E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \lambda} \right|_{\lambda=\lambda^*} = 0 \quad (3.3)$$

para todo valor \mathbf{B} y $\{\mathbf{M}_k\}$, donde λ^* representan los valores óptimos de λ . Esto a la vez determina la dependencia de λ^* con \mathbf{B} y $\{\mathbf{M}_k\}$.

Considerando a la energía como la funcional anterior, para el caso de funciones de onda variacionales, se tiene luego de aplicar la regla de la cadena

$$\frac{dE(\mathbf{B}, \mathbf{M}_k; \lambda)}{d\mathbf{M}_k} = \left[\frac{\partial E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \mathbf{M}_k} + \frac{\partial E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \lambda} \frac{\partial \lambda}{\partial \mathbf{M}_k} \right] \Big|_{\lambda=\lambda^*}$$

por lo que

$$\frac{dE(\mathbf{B}, \mathbf{M}_k; \lambda^*)}{d\mathbf{M}_k} = \frac{\partial E(\mathbf{B}, \mathbf{M}_k; \lambda^*)}{\partial \mathbf{M}_k}$$

Las correspondientes expresiones para las constantes de apantallamiento pueden ser obtenidas en forma casi directa

$$\begin{aligned}\frac{d^2 E(\mathbf{B}, \mathbf{M}_k; \lambda^*)}{d\mathbf{B} d\mathbf{M}_k} &= \left[\left(\frac{\partial}{\partial \mathbf{B}} + \frac{\partial \lambda}{\partial \mathbf{B}} \frac{\partial}{\partial \lambda} \right) \frac{\partial E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \mathbf{M}_k} \right] \Big|_{\lambda=\lambda^*} \\ &= \left[\frac{\partial^2 E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \mathbf{B} \partial \mathbf{M}_k} + \frac{\partial^2 E(\mathbf{B}, \mathbf{M}_k; \lambda)}{\partial \mathbf{M}_k \partial \lambda} \frac{\partial \lambda}{\partial \mathbf{B}} \right] \Big|_{\lambda=\lambda^*}\end{aligned}$$

Si $\mathbf{B} = \mathbf{M}_k = \mathbf{0}$ ⁽⁵⁾ las expresiones toman la misma forma. Por simplicidad de notación de aquí en más se asumirá esta condición y también $\lambda = \lambda^*$. Las constantes de apantallamiento del núcleo K satisfacen entonces

$$\sigma_K = \mathbf{1} + \frac{\partial^2 E}{\partial \mathbf{B} \partial \mathbf{M}_K} + \frac{\partial^2 E}{\partial \mathbf{M}_K \partial \lambda} \frac{\partial \lambda}{\partial \mathbf{B}} \quad (3.4)$$

Es importante notar que, para el cálculo de los apantallamientos con funciones de onda variacionales, sólo las funciones de respuesta de primer orden (las derivadas de λ) respecto de \mathbf{B} son necesarias.

Dado que

$$\frac{d^2 E(\mathbf{B}, \mathbf{M}_k; \lambda^*)}{d\mathbf{B} d\mathbf{M}_k} = \frac{d^2 E(\mathbf{B}, \mathbf{M}_k; \lambda^*)}{d\mathbf{M}_k d\mathbf{B}}$$

podría haberse procedido en términos de \mathbf{M}_K . Es más ventajosa la forma anterior porque la misma sólo requiere tres funciones de respuesta. La segunda opción requiere tres funciones de respuesta por cada núcleo.

No es posible aprovechar la ventaja anterior para el cálculo de las constantes de acoplamiento espín-espín reducidas. A estas se las puede expresar en términos de las funciones de respuesta según

$$\mathbf{K}_{KL} = \frac{\partial^2 E}{\partial \mathbf{M}_K \partial \mathbf{M}_L} + \frac{\partial^2 E}{\partial \mathbf{M}_K \partial \lambda} \frac{\partial \lambda}{\partial \mathbf{M}_L} - \mathbf{D}_{KL} \quad (3.5)$$

A efectos prácticos esto significa que el costo de cálculo relativo entre las constantes de acoplamiento y las de apantallamiento aumenta con el número de núcleos de la molécula.

Una forma de evaluar las funciones de respuesta puede obtenerse al diferenciar la ecuación 3.3. La forma funcional es independiente de la variable elegida, aquí se lo mostrará para \mathbf{B} ,

$$\frac{d}{d\mathbf{B}} \frac{\partial E}{\partial \lambda} = \frac{\partial^2 E}{\partial \mathbf{B} \partial \lambda} + \frac{\partial^2 E}{\partial \lambda^2} \frac{\partial \lambda}{\partial \mathbf{B}}$$

⁵Ese es el caso útil para obtener los apantallamientos y constantes de acoplamiento.

que puede ser reescrito como

$$\frac{\partial^2 E}{\partial \lambda^2} \frac{\partial \lambda}{\partial \mathbf{B}} = -\frac{\partial^2 E}{\partial \lambda \partial \mathbf{B}} \quad (3.6)$$

3.4.2. Hamiltoniano electrónico en segunda cuantización

Los tratamientos modernos del cálculo de las propiedades magnéticas están enmarcados en el formalismo de segunda cuantización. Con él, puede expresarse al hamiltoniano electrónico en términos de los operadores de excitación singlete y triplete según

$$\begin{aligned} H(\mathbf{B}, \mathbf{M}) &= H^{(0)} + \mathbf{B}^T (\mathbf{h}_B^{orb} + \mathbf{h}_B^{spn}) + \sum_K \mathbf{M}_K^T (\mathbf{h}_K^{pso} + \mathbf{h}_K^{fc} + \mathbf{h}_K^{sd}) + \sum_K \mathbf{B}^T \mathbf{h}_{BK}^{dia} \mathbf{M}_K \\ &\quad + \sum_{K>L} \mathbf{M}_K^T \mathbf{h}_{KL}^{dso} \mathbf{M}_L + \mathbf{B}^T \mathbf{h}_{BB}^{dia} \mathbf{B} - \mathbf{B}^T \sum_K \mathbf{M}_K + \sum_{K>L} \mathbf{M}_K^T \mathbf{D}_{KL} \mathbf{M}_L \\ H^{(0)} &:= \sum_{pq} h_{pq}(\mathbf{B}) E_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs}(\mathbf{B}) (E_{pq} E_{rs} - \delta_{rq} E_{ps}) + \sum_{K>L} \frac{Z_K Z_L}{R_{KL}} \\ h_{pq} &:= \langle \phi_p(\mathbf{B}) | -\frac{1}{2} \nabla^2 - \sum_K \frac{Z_K}{r_K} | \phi_q(\mathbf{B}) \rangle \\ g_{pqrs} &:= \langle \phi_p(\mathbf{r}_1, \mathbf{B}) \phi_q(\mathbf{r}_1, \mathbf{B}) | \frac{1}{r_{12}} | \phi_r(\mathbf{r}_2, \mathbf{B}) \phi_s(\mathbf{r}_2, \mathbf{B}) \rangle \\ \mathbf{h}_B^{orb} &:= -\frac{1}{2} i \sum_{pq} \langle \psi_p(\mathbf{B}) | \mathbf{r}_O \times \nabla | \psi_q(\mathbf{B}) \rangle E_{pq} \\ \mathbf{h}_K^{pso} &:= -i \alpha^2 \sum_{pq} \langle \psi_p(\mathbf{B}) | \frac{\mathbf{r}_K \times \nabla}{r_K^3} | \psi_q(\mathbf{B}) \rangle E_{pq} \\ \mathbf{h}_B^{spn} &:= \sum_p \mathbf{T}_{pp} \\ \mathbf{h}_K^{fc} &:= \frac{8\pi\alpha^2}{3} \sum_{pq} \langle \psi_p(\mathbf{B}) | \delta(r_K) | \psi_q(\mathbf{B}) \rangle \mathbf{T}_{pq} \\ \mathbf{h}_K^{sd} &:= -\alpha^2 \sum_{pq} \langle \psi_p(\mathbf{B}) | \frac{r_K^2 \mathbf{1} - 3\mathbf{r}_K \mathbf{r}_K^T}{r_K^5} | \psi_q(\mathbf{B}) \rangle \mathbf{T}_{pq} \end{aligned}$$

$$\begin{aligned}
\mathbf{h}_{BB}^{dia} &:= \frac{1}{8} \sum_{pq} \langle \psi_p(\mathbf{B}) | r_O^2 \mathbf{1} - \mathbf{r}_O \mathbf{r}_O^T | \psi_q(\mathbf{B}) \rangle E_{pq} \\
\mathbf{h}_{BK}^{dia} &:= \frac{\alpha^2}{2} \sum_{pq} \langle \psi_p(\mathbf{B}) | \frac{(\mathbf{r}_O^T \mathbf{r}_K)^2 \mathbf{1} - \mathbf{r}_O \mathbf{r}_K^T}{r_K^3} | \psi_q(\mathbf{B}) \rangle E_{pq} \\
\mathbf{h}_{KL}^{dso} &:= \frac{\alpha^4}{2} \sum_{pq} \langle \psi_p(\mathbf{B}) | \frac{(\mathbf{r}_K^T \mathbf{r}_L)^2 \mathbf{1} - \mathbf{r}_K \mathbf{r}_L^T}{r_K^3 r_L^3} | \psi_q(\mathbf{B}) \rangle E_{pq}
\end{aligned}$$

E_{pq} y \mathbf{T}_{pq} representan a los operadores de excitación singlete y triplete, respectivamente. Para más detalles ver Apéndice A.

Con el método de Hartree-Fock en su variación restringida, se representa a un dado estado de capa cerrada por un determinante de Slater $|\kappa\rangle$

$$|\kappa\rangle = \exp(-\hat{\kappa}) \left(\prod_i a_{i\alpha}^\dagger a_{i\beta}^\dagger \right) |\text{vac}\rangle$$

El vector $|\text{vac}\rangle$ representa el estado vacío, y los $a_{i\gamma}$ representan operadores de creación⁶.

El operador de rotación ($\hat{\kappa}$) puede expresarse como una combinación lineal de operadores de rotaciones elementales

$$|\hat{\kappa}\rangle = \sum_i \hat{\kappa}_i |\kappa_i\rangle$$

que no dependen del campo magnético para propiedades magnéticas independientes del tiempo[14]. En este marco, el método de Hartree-Fock consiste en la minimización variacional de la energía respecto de los parámetros de rotación.

$$E_{HF}^0 = \min_{\kappa} \langle \kappa | \hat{H} | \kappa \rangle$$

El operador de rotación en términos de los operadores de excitación (ec. A.4),

$$\hat{\kappa} = i \sum_{p>q} {}^I \kappa_{pq}^0 + \sum_{p>q} {}^R \kappa_{pq}^0 E_{pq}^- + i \sum_{p>q} \alpha \sum_{p>q} {}^I \kappa_{pq}^\alpha T_{pq}^{\alpha+} + \sum_{\alpha} \sum_{p>q} {}^R \kappa_{pq}^\alpha T_{pq}^{\alpha-}$$

es usado en presencia de perturbaciones magnéticas⁷, y es más completo que el usado en ausencia de ellas. El último sólo contiene rotaciones reales singletes. Aún así, no todos los

⁶Esto está descrito en el Apéndice A

⁷Campo magnético externo.

términos de la ecuación anterior son necesarios. Sólo se necesitan rotaciones que vinculen a los orbitales ocupados con los desocupados.

Como en primera cuantización, el método de Hartree-Fock puede reducirse a la resolución de las ecuaciones de Fock. Estas son ecuaciones monoelectrónicas, que consisten en un problema de autovalores y autovectores para el operador de Fock (\hat{f}). Este es un hamiltoniano efectivo que representa a un electrón inmerso en un campo, donde el efecto de los otros electrones sobre el primero es considerado en una forma promediada. La parte monoelectrónica (h) es tomada del hamiltoniano exacto⁸. La interacción promediada es considerada en un potencial efectivo \hat{V}

$$\begin{aligned}\hat{f} &:= h + \hat{V} \\ \hat{V} &:= \sum_{pq} V_{pq} E_{pq} \\ V_{pq} &:= \sum_i \sum_i (2g_{pqii} - g_{piii})\end{aligned}$$

En breve, el método de Hartree-Fock permite encontrar variacionalmente un conjunto de orbitales, para construir un determinante de Slater que aproximadamente representa a la función de onda, que se obtienen como autofunciones de la del operador de Fock

$$\hat{f} a_{p\sigma}^\dagger |\text{vac}\rangle = \epsilon_p a_{p\sigma}^\dagger |\text{vac}\rangle$$

La energía puede expandirse en series de alguna perturbación⁹. Puede demostrarse que las siguientes expresiones son correctas para dicha expansión

$$\begin{aligned}E_{HF}^{(0)} &= \langle HF | \hat{H}^{(0)} | HF \rangle \\ E_{HF}^{(1)} &= \langle HF | \hat{H}^{(1)} | HF \rangle \\ E_{HF}^{(2)} &= \langle HF | \hat{H}^{(2)} | HF \rangle + \langle HF | [\hat{\kappa}^{(1)}, \hat{H}^{(1)}] | HF \rangle\end{aligned}$$

donde $|HF\rangle$ representa al vector de estado obtenido mediante el método de Hartree-Fock.

⁸Incluye a la energía cinética y las atracciones con los núcleos

⁹Como se comentó al inicio de la sección anterior.

El término de primer orden en la energía es nulo. Esto es así porque, para especies de capa cerrada, las derivadas del hamiltoniano respecto del campo magnético y los momentos magnéticos son nulas. Para el campo magnético, $dH/d\mathbf{B} = \mathbf{h}_B^{orb} + \mathbf{h}_B^{spn}$. Al aplicarlos a un estado cualquiera ($|0\rangle$) de capa cerrada,

$$\mathbf{h}_B^{spn}|0\rangle = \sum_i \mathbf{m}_i|0\rangle = -\sum_i \mathbf{s}_i|0\rangle = \mathbf{S}|0\rangle = \mathbf{0}$$

mientras que $\mathbf{h}_B^{orb}|0\rangle$ genera un singlete imaginario. En el caso de los momentos magnéticos nucleares

$$dH/d\mathbf{M}_K = \mathbf{h}_K^{pso} + \mathbf{h}_K^{sd} + \mathbf{h}_K^{fc}$$

al aplicarlos a $|0\rangle$ estos tres operadores dan estados (de izquierda a derecha): singlete imaginario, triplete real, triplete real.

Los operadores de espín triplete no contribuyen porque su efecto es generar una función de onda de espín triplete¹⁰. Los valores esperados de operadores hermíticos imaginarios son nulos en estados que representen sistemas de capa cerrada.

Para calcular la corrección de segundo orden a la energía se necesita la función de onda perturbada hasta primer orden, la cual es determinada por la corrección por perturbación del operador de rotación ($\hat{\kappa}^{(1)}$). Este es determinado por las ecuaciones

$$0 = \langle HF | [\hat{\kappa}_i, \hat{H}^{(0)}] | HF \rangle - \sum_j \langle HF | [\hat{\kappa}_i, [\hat{\kappa}_j, \hat{H}^{(0)}]] | HF \rangle \hat{\kappa}_j^{(1)} = \langle HF | [\hat{\kappa}_i, \hat{H}^{(1)}] | HF \rangle$$

El primer término de la corrección de segundo orden a la energía, corresponde a la sumatoria sobre los distintos momentos magnéticos nucleares, de los valores esperados de las derivadas de segundo orden del hamiltoniano

$$\begin{aligned} \langle HF | \frac{d^2 \hat{H}}{d\mathbf{B} d\mathbf{M}_K} | HF \rangle &= \langle HF | \mathbf{h}_{BK}^{dia} + \frac{d\mathbf{h}_K^{pso}}{d\mathbf{B}} | HF \rangle - \mathbf{1} \\ \langle HF | \frac{d^2 \hat{H}}{d\mathbf{M}_K d\mathbf{M}_L} | HF \rangle &= \langle HF | \mathbf{h}_{KL}^{dso} | HF \rangle + \mathbf{D}_{KL} \end{aligned}$$

¹⁰ En casos de capa abierta resultan la principal contribución

No se incluyen términos de segundo orden respecto del campo magnético. Esto es porque son nulos como puede verse por simple inspección del hamiltoniano 3.2. La contribución de $\frac{d\mathbf{h}_K^{ps0}}{d\mathbf{B}}$ se debe a la dependencia de los orbitales con el campo magnético.

El segundo término de la corrección de segundo orden a la energía es

$$\langle HF | [\hat{\kappa}_i, \frac{d\hat{H}}{d\mathbf{B}}] | HF \rangle \begin{cases} \langle HF | [iE_{pq}^+, \mathbf{h}_B^{orb} + \frac{d\hat{H}^{(0)}}{d\mathbf{B}}] | HF \rangle & \text{si } \hat{\kappa}_i = iE_{pq}^+ \\ 0 & \text{otro caso} \end{cases}$$

$$\langle HF | [\hat{\kappa}_i, \frac{d\hat{H}}{d\mathbf{M}_K}] | HF \rangle \begin{cases} \langle HF | [iE_{pq}^+, \mathbf{h}_K^{ps0}] | HF \rangle & \text{si } \hat{\kappa}_i = iE_{pq}^+ \\ \langle HF | [T_{pq}^{\alpha-}, \mathbf{h}_K^{fc} + \mathbf{h}_K^{sd}] | HF \rangle & \text{si } \hat{\kappa}_i = T_{pq}^{\alpha-} \\ 0 & \text{en otro caso} \end{cases}$$

A partir de estas ecuaciones se pueden obtener formas explícitas para las constantes de apantallamiento y acoplamiento indirecto en términos de la teoría de perturbación variacional (ecuaciones 3.4 y 3.5)

$$\begin{aligned} \sigma_K &= \langle HF | \mathbf{h}_{BK}^{dia} + \frac{d\mathbf{h}_K^{ps0}}{d\mathbf{B}} | HF \rangle + \sum_{p>q} \langle HF | [iE_{pq}^+, (\mathbf{h}_K^{ps0})^T] | HF \rangle \frac{d^I \kappa_{pq}^0}{d\mathbf{B}} \\ \mathbf{K}_{KL} &= \langle HF | \mathbf{h}_{KL}^{dso} | HF \rangle + \sum_{p>q} \langle HF | [iE_{pq}^+, (\mathbf{h}_K^{ps0})^T] | HF \rangle \frac{d^I \kappa_{pq}^0}{d\mathbf{M}_L} \\ &\quad + \sum_{\alpha} \sum_{p>q} \langle HF | [T_{pq}^{\alpha-}, (\mathbf{h}_K^{fc})^T + (\mathbf{h}_K^{sd})^T] | HF \rangle \frac{d^R \kappa_{pq}^{\alpha}}{d\mathbf{M}_L} \end{aligned}$$

Estas expresiones contienen a tres funciones de respuesta (ver ec. 3.6), cuya forma explícita es

$$\begin{aligned} \sum_{r>s} \langle HF | [iE_{pq}^+, [iE_{rs}^+, \hat{H}]] | HF \rangle \frac{d^I \kappa_{rs}^0}{d\mathbf{B}} &= -\langle HF | [iE_{pq}^+, \mathbf{h}_B^{orb} + \frac{d\hat{H}^{(0)}}{d\mathbf{B}}] | HF \rangle \\ \sum_{r>s} \langle HF | [iE_{pq}^+, [iE_{rs}^+, \hat{H}]] | HF \rangle \frac{d^I \kappa_{rs}^0}{d\mathbf{M}_K} &= -\langle HF | [iE_{pq}^+, \mathbf{h}_K^{ps0}] | HF \rangle \\ \sum_{r>s} \langle HF | [T_{pq}^{\alpha-}, [[T_{rs}^{\alpha-}, \hat{H}]] | HF \rangle \frac{d^R \kappa_{rs}^0}{d\mathbf{M}_K} &= -\langle HF | [T_{pq}^{\alpha-}, \mathbf{h}_K^{fc} + \mathbf{h}_K^{sd}] | HF \rangle \end{aligned}$$

3.5. El problema del origen para el potencial vector magnético

Como fue mencionado, una forma válida para el potencial vector magnético es

$$\mathbf{A}_O(\mathbf{r}) = \frac{1}{2} \mathbf{B} \times (\mathbf{r} - \mathbf{O})$$

Para cualquier origen (\mathbf{O}) se satisface la definición del potencial vector magnético. Sin embargo, la forma del hamiltoniano varía con \mathbf{O} . Naturalmente los observables físicos no son dependientes de este origen. Dos potenciales están vinculados por una transformación de *gauge*¹¹, por lo que se puede usar estas transformaciones para obtener un nuevo potencial con origen en \mathbf{O}'

$$\mathbf{A}_{O'}(\mathbf{r}) = \mathbf{A}_O(\mathbf{r}) + \nabla f(\mathbf{r}) \quad (3.7)$$

Con tal fin se define a la función f

$$f(\mathbf{r}) := \frac{1}{2} \mathbf{B} \times (\mathbf{O} - \mathbf{O}') \cdot \mathbf{r}$$

Puede demostrarse que la forma general para la transformación de la función de onda correspondiente a una transformación del potencial vector magnético (ec. 3.7) es

$$\psi(\mathbf{r})' = e^{if(\mathbf{r})} \psi(\mathbf{r})$$

Entonces, para el caso del cambio de origen

$$\psi'(\mathbf{r}) = e^{i \frac{1}{2} \mathbf{B} \times (\mathbf{O} - \mathbf{O}') \cdot \mathbf{r}} \psi(\mathbf{r})$$

La función de onda sólo se ve afectada en un factor de fase, lo cual evidencia la ausencia de efectos sobre los observables físicos

$$\langle \psi'(\mathbf{r}) | \mathcal{X} | \psi'(\mathbf{r}) \rangle = \langle \psi(\mathbf{r}) | \mathcal{X} | \psi(\mathbf{r}) \rangle$$

En tratamientos aproximados, donde el espacio variacional no incluye a las funciones de onda exactas, lo anterior no es estrictamente cierto. La elección del origen deja de ser

¹¹Esto fue discutido en el capítulo de mecánica clásica.

arbitraria y debe realizarse con buen criterio. En algunos casos existe un *origen natural*[15] que resulta conveniente. Esto es apreciable a partir de teoría de perturbaciones estacionaria¹².

Sea por caso el átomo de hidrógeno inmerso en un campo magnético estático y uniforme, su hamiltoniano es

$$H = \frac{1}{2}p^2 + \frac{e}{2mc}\mathbf{A} \cdot \mathbf{p} + \frac{e}{2mc^2}A^2 + V(\mathbf{r})$$

el cual puede reescribirse, mediante el reemplazo del potencial vector magnético, como

$$H = \frac{1}{2}p^2 + \frac{e}{2mc}\mathbf{B} \cdot \mathbf{l} - \frac{e}{2mc}\mathbf{B} \times \mathbf{O} \cdot \mathbf{p} + \frac{e^2}{8mc^2} \left(B^2(\mathbf{r} - \mathbf{O})^2 - [\mathbf{B} \cdot (\mathbf{r} - \mathbf{O})]^2 \right) + V(\mathbf{r})$$

Para simplificar el tratamiento, se supondrá que la única componente no nula del campo corresponde al eje z

$$H = \frac{1}{2}p^2 + \frac{e}{2mc}Bl_z - \frac{eB}{2mc}(Xp_x - Yp_y) + \frac{e^2B^2}{8mc^2}[(x - X)^2 + (y - Y)^2] + V(\mathbf{r})$$

Según la teoría de perturbaciones estacionaria, para un hamiltoniano H , expresado como un hamiltoniano no perturbado H_0 sobre el cual se ejerce una perturbación general del tipo

$$H = H_0 + \sum_i \lambda^i V_i$$

las primeras correcciones a la energía son

$$E_1 = \langle \psi_0 | V_1 | \psi_0 \rangle$$

$$E_2 = \langle \psi_0 | V_2 | \psi_0 \rangle + \langle \psi_0 | V_1 | \psi_1 \rangle$$

Limitando el tratamiento a estados correspondientes a orbitales s , se obtiene que E_1 es nula para el caso de interés. Esto es así cualquiera sea el origen $\mathbf{O} = (X, Y, Z)$ elegido.

La expresión para E_2 depende del origen escogido. Escogiendo la posición del núcleo como origen de coordenadas ($\mathbf{O} = (0, 0, 0)$), resulta que

¹²Puede encontrarse una discusión muy completa del tema en la referencia [15]

$$E_2 = \frac{e^2}{mc^2} \langle \psi_0 | x^2 + y^2 | \psi_0 \rangle + \frac{e}{2mc} \langle \psi_0 | l_x | \psi_1 \rangle$$

Al primer término se lo llama término diamagnético o de Langevin. Al segundo se lo llama paramagnético o de van Vleck. En estas condiciones el término de van Vleck es nulo. Usando un origen arbitrario, el resultado obtenido es

$$E_2 = \frac{e^2}{mc^2} \langle \psi_0 | x^2 + y^2 | \psi_0 \rangle + \frac{e}{2mc} \langle \psi_0 | l_x | \psi_1 \rangle + \frac{e^2 X^2}{8mc^2} - \frac{eX}{2mc} \langle \psi_0 | p_y | \tilde{\psi}_1 \rangle$$

Puede demostrarse que los dos últimos términos se cancelan entre si y sólo si ¹³

$$[p_y, y] = \frac{\hbar}{i}$$

Esta relación es válida para funciones de onda exactas, pero no para sus representaciones en bases finitas¹⁴. Por lo tanto, el origen natural es preferible porque es exacto hasta un mayor orden, en el marco de la teoría de perturbaciones estacionaria, para funciones de onda aproximadas.

Entonces, existe un origen natural para los sistemas monoeléctricos de simetría esférica¹⁵, y corresponde al respectivo núcleo. En tales casos puede incluirse esta corrección al factor de fase de la función de onda, para obtener soluciones más precisas.

En el caso de moléculas no existe un origen natural. Existen varios métodos para tratar este problema. Una solución elegante es acoplar el factor de fase a los orbitales atómicos, de modo que puede tenerse un origen diferente para cada núcleo. Por ejemplo, un orbital de tipo GTO (spherical harmonic Gaussian Type Orbital) tendría la forma

$$\phi(\mathbf{r}, \mathbf{B}) = e^{[-\frac{1}{2}i\mathbf{B} \times (\mathbf{N}-\mathbf{O}) \cdot \mathbf{r}]} S_{lm}(\mathbf{r}) e^{-ar^2}$$

donde \mathbf{N} es el vector que representa la posición del núcleo en cuestión.

Estos orbitales son llamados *orbitales atómicos de London*¹⁶ o *Gauge Invariant Atomic Orbitals* (GIAO). Los cálculos realizados para esta tesis utilizan este tratamiento del origen del potencial vector magnético.

¹³ Lo siguiente es requerido para obtener a $\tilde{\psi}_1$ en forma exacta.

¹⁴ Salvo que casualmente $p_y \phi$ y $y \phi$ estén representadas en forma exacta.

¹⁵ Esto puede extenderse a átomos en general

¹⁶ Fueron utilizados por primera vez por London en 1937.

3.6. Ideas principales sobre el capítulo

En este capítulo se presentó al hamiltoniano efectivo de espín. Este es un hamiltoniano aproximado utilizado para la interpretación de espectros de RMN. Es útil gracias a sus parámetros *ad hoc* que son ajustados experimentalmente. Se discutió la interpretación fenomenológica de sus distintos términos. Se mostró la simplificación del hamiltoniano que surge para moléculas en libre rotación, la cual es de gran importancia en fase gaseosa. Se propuso una demostración de la misma.

Dado que este hamiltoniano no proviene de primeros principios, la predicción de sus parámetros no es directa. Se mostró el vínculo entre ambos hamiltonianos a partir de desarrollo en series de Taylor de la energía. Se mostró que las derivadas presentes en la expansión de series de potencias permiten la obtención de los parámetros *ad hoc*. Se mostró la forma explícita de esos parámetros de acuerdo a las expresiones de Ramsey. Esto permite comprender el significado de las matrices usadas para la construcción de los descriptores propuestos en esta tesis. Contrariamente, dichas expresiones no permiten el cálculo por cuestiones de tipo prácticas.

Se mostró los lineamientos generales para el cálculo de las propiedades magnéticas en cuestión. Particularmente para el caso de métodos variacionales. Esto fue enmarcado en la teoría de perturbaciones variacional, la cual fue brevemente descrita. Se vinculó al conjunto de parámetros, que en ella aparece, con los coeficientes de una combinación lineal de operadores elementales de rotación. Se describió brevemente al modelo de Hartree-Fock en el formalismo de segunda cuantización (el cual se desarrolló con mayor detalle en el apéndice). En este marco, el problema puede entenderse como la obtención de las rotaciones óptimas (en sentido variacional), y de ahí surge el vínculo con los parámetros de la teoría de perturbaciones variacional. Finalmente se resumieron las expresiones explícitas, para la obtención de las constantes de apantallamiento y de acoplamiento indirecto, mediante el método de Hartree-Fock en el formalismo de segunda cuantización.

Finalmente se describió al problema de la elección de la función *gauge*, y al método GIAO.

Capítulo 4

Teoría del Funcional de la Densidad

4.1. Introducción

El hamiltoniano (\mathcal{H}) de un sistema de N electrones que interactúan entre sí, inmersos en un campo de potencial externo, puede ser expresado como

$$\mathcal{H} = \mathcal{T} + \mathcal{W} + \sum_i^N \mathcal{V}(\mathbf{r}_i)$$

donde los operadores energía cinética ($\mathcal{T} = -\frac{1}{2} \sum_i \nabla_i^2$) y repulsión electrónica ($\mathcal{W} = \sum_{i>j} r_{ij}^{-1}$) son iguales para cualquier sistema de N electrones, mientras que \mathcal{V} no. Los potenciales de mayor interés son aquellos capaces de mantener unidos a los N electrones en un estado fundamental. En particular interesan aquellos de tipo coulombico tales como los generados por los núcleos.

La densidad electrónica es definida con

$$\rho(\mathbf{r}) = N \int \cdots \int d\sigma_1 d\mathbf{x}_2 \dots d\mathbf{x}_N |\Psi(\mathbf{x}, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2$$

donde \mathbf{x}_i representa las variables electrónicas de espacio y espín del electrón i , y σ_i su variable de espín.

Si el estado de uno de estos sistemas es $|\Psi\rangle$, el valor esperado para su energía es

$$E = \langle \Psi | \mathcal{H} | \Psi \rangle = \langle \Psi | \mathcal{T} + \mathcal{W} | \Psi \rangle + \langle \Psi | \sum_i^N \mathcal{V}(\mathbf{r}_i) | \Psi \rangle$$

Dado que los potenciales son multiplicativos, se cumple

$$E = \langle \Psi | \mathcal{T} + \mathcal{W} | \Psi \rangle + \langle \mathcal{V} | \rho \rangle$$

donde $\mathcal{V}(\mathbf{r}) := \sum_i^N \mathcal{V}(\mathbf{r}_i)$.

Con un poco de álgebra puede demostrarse que, funciones de onda idénticas (a menos un factor de fase) de estados fundamentales corresponden a potenciales idénticos (a menos una constante aditiva). Además, que potenciales que se diferencien en más que una constante, dan origen a funciones de onda que se diferencian en más que un factor de fase.

Puede demostrarse el siguiente teorema[16] haciendo uso del principio variacional de Rayleigh-Ritz,

Teorema 4.1.1 (Hohenberg-Kohn (HK)). *Cada densidad correspondiente a un estado fundamental de un sistema N electrónico, es determinada por uno y sólo un potencial (salvo una constante aditiva).*

A estas densidades se las clasifica como \mathcal{V} -representables. Se denotará por \mathbb{A}_N al conjunto de dichas densidades para sistemas de N electrones. Una densidad $\rho \in \mathbb{A}_N$ determina al potencial $\mathcal{V}_\rho + c$ y consecuentemente a la función $\gamma\Psi_\rho$. Esto es interesante dado que la densidad electrónica es función de \mathbf{r} (y eventualmente del tiempo) mientras que la función de onda lo es de los elementos de $\{\mathbf{r}_i, \sigma_i\}$ (y eventualmente del tiempo). Nótese que \mathcal{V} determina las densidades correspondientes a un conjunto de estados degenerados.

Se define al funcional de Hohenberg-Kohn con

$$\mathcal{F}_{HK}[\rho] := \langle \Psi_\rho | \mathcal{T} + \mathcal{W} | \Psi_\rho \rangle = E[\mathcal{V}_\rho] - \langle \mathcal{V}_\rho | \rho \rangle$$

siendo $\rho \in \mathbb{A}_N$ verdadero.

Comentario 4.1.2. Nótese que se está definiendo al funcional de Hohenberg-Kohn según el resultado de su aplicación sobre las densidades pertenecientes a \mathbb{A}_N , y que en la definición no se explicita su forma. En forma análoga se definen las funcionales

$$\begin{aligned} \mathcal{T}[\rho] &:= \langle \Psi_\rho | \mathcal{T} | \Psi_\rho \rangle \\ \mathcal{W}[\rho] &:= \langle \Psi_\rho | \mathcal{W} | \Psi_\rho \rangle \end{aligned}$$

Se utilizó al mismo símbolo para estas funcionales y sus correspondientes operadores. Su distinción debería ser clara por contexto.

Aplicando el principio variacional puede demostrarse que, para un potencial arbitrario que dé origen a un estado fundamental, se cumple

$$E[\mathcal{V}] \leq \mathcal{F}_{HK}[\rho] + \langle \mathcal{V} | \rho \rangle$$

Dicho conjunto de potenciales se denotará por \mathbb{V}_N . En principio se podría obtener la energía correspondiente a un $\mathcal{V} \in \mathbb{V}_N$ realizando variaciones sobre las densidades $\{\rho | \rho \in \mathbb{A}_N\}$ hasta obtener una densidad tal que

$$E[\mathcal{V}] = \min_{\rho \in \mathbb{A}_N} (\mathcal{F}_{HK}[\rho] + \langle \mathcal{V} | \rho \rangle)$$

Esto conlleva dificultades dado que no se conoce la forma explícita de \mathcal{F}_{HK} , y los conjuntos \mathbb{A}_N y \mathbb{V}_N son difíciles de caracterizar. Es decir, si bien estos conjuntos están claramente definidos, no es fácil determinar si efectivamente una dada densidad pertenece a \mathbb{A}_N .

Denotando con \mathcal{L}^p al espacio de Lebesgue de las funciones p -integrables, definidas como aquellas cuya norma satisface

$$\|f\|_p = \left(\int d\mathbf{r} |f(\mathbf{r})|^p \right)^{1/p} < \infty$$

y considerando que la densidad electrónica debe ser integrable

$$\int d\mathbf{r} \rho(\mathbf{r}) = N$$

se concluye que $\rho \in \mathcal{L}^1$. Puede probarse además que la finitud de la energía cinética implica que $\rho \in \mathcal{L}^3$ ⁽¹⁾. Se concluye entonces que $\rho \in \mathcal{L}^1 \cap \mathcal{L}^3$. Los potenciales pertenecen al espacio (de Banach) dual de aquel de las densidades, del primero puede demostrarse que es $\mathcal{L}^{\frac{3}{2}} + \mathcal{L}^\infty$. El subespacio de $\mathcal{L}^1 \cap \mathcal{L}^3$ tal que ρ proviene de una función de onda antisimétrica normalizada para sistemas de N partículas es convexo y resulta menos restrictivo que \mathbb{A}_N . Para densidades de dicho espacio se define

$$\mathcal{F}_{LL}[\rho] := \langle \Psi_\rho | \mathcal{T} + \mathcal{W} | \Psi_\rho \rangle$$

por lo que si además si $\rho \in \mathbb{A}_N$

$$\mathcal{F}_{LL}[\rho] = \mathcal{F}_{HK}[\rho]$$

¹La demostración no es obvia. Ver por ejemplo [17].

Puede demostrarse que esta funcional no es convexa, lo cual es inconveniente porque la búsqueda del mínimo de energía puede llevar a otros puntos estacionarios.

Definición 4.1.3. Sea \mathcal{X} un espacio real normado, \mathcal{X}^* el espacio dual de \mathcal{X} , $\langle \cdot, \cdot \rangle : \mathcal{X}^* \times \mathcal{X} \rightarrow \mathbb{R}$, $f : \mathcal{X} \rightarrow \mathbb{R} \cup \infty$, $f^* : \mathcal{X}^* \rightarrow \mathbb{R} \cup \infty$, se define a la **transformada de Legendre-Frenchel o conjugada convexa** de f , f^* , con

$$f^*(x^*) := \sup\{\langle x^*, x \rangle - f(x) \mid x \in \mathcal{X} \text{ y } x^* \in \mathcal{X}^*\}$$

Para forzar la convexidad de la funcional, se utiliza la funcional convexa bi-conjugada de \mathcal{F}_{LL} (funcional de Lieb). Esto significa aplicar dos veces la transformación de Legendre sobre \mathcal{F}_{LL} . Con ello se obtiene la envolvente convexa de la primera funcional, por lo que el supremo de ambas funcionales es el mismo.

4.2. Forma explícita de las funcionales

Se mostró que existen funcionales que permiten encontrar el valor exacto de la energía de los sistemas a partir de sus densidades. Algunas de sus propiedades matemáticas fueron encontradas, pero su forma funcional es desconocida. En la práctica es necesario establecer alguna forma funcional.

4.2.1. Modelo de Thomas-Fermi-Dirac

Una forma explícita pero inexacta de las funcionales corresponde a las del modelo de Thomas-Fermi[18, 19]. Se llama gas de Fermi a un conjunto hipotético de fermiones que no interactúan coulómbicamente entre sí. Si dicho gas se encuentra confinado en una caja, cuyo potencial es constante dentro de ella pero infinito en su exterior, se puede utilizar las soluciones de la partícula en la caja para encontrar su energía.

Para una partícula sujeta a estas condiciones la ecuación de Schrödinger tendrá la forma

$$-\frac{1}{2}\nabla^2\psi(\mathbf{r}) = E\psi(\mathbf{r})$$

y las energías permitidas serán $E_k = k^2/2$, con $\mathbf{k} := 2\pi l^{-1}(n_x, n_y, n_z)$ y $n_i \in \mathbb{Z}$.

La energía del gas de Fermi se puede minimizar, satisfaciendo el postulado de simetría, permitiendo que hasta un máximo de dos electrones compartan los tres números cuánticos. Para dicha energía, $\sqrt{n_x^2 + n_y^2 + n_z^2}$ es aproximadamente constante. Entonces los posibles

estados pueden ser graficados en una esfera de radio constante, que implica un máximo valor de k llamado radio de Fermi (k_F).

Como las partículas son independientes,

$$\rho(\mathbf{r}) = 2 \sum |\psi(\mathbf{r})|^2$$

Dentro de la caja las funciones de onda son

$$\psi(\mathbf{r}) = l^{-\frac{3}{2}} e^{i\mathbf{k}\cdot\mathbf{r}}$$

Si la caja es lo suficientemente grande es posible reemplazar a las sumas por integrales. Haciendo los reemplazos $2\pi l^{-1}(n_x, n_y, n_z) = (k_x, k_y, k_z)$ y $d\mathbf{n} = l^3/(8\pi^3)d\mathbf{k}$ pueden obtenerse expresiones para la energía y para la densidad en función de k_F . La energía puede expresarse entonces como una función de la densidad. El resultado que se obtiene es

$$E = C_F \rho^{5/3}$$

con $C_F := 3/10 (3\pi^2)^{2/3}$.

Nótese que esta funcional se refiere esencialmente a energía cinética. La funcional de Thomas-Fermi \mathcal{F}_{TF}

$$\mathcal{F}_{TF} := \mathcal{T}_{TF} + \mathcal{J}_{TF} \quad (4.1)$$

$$\mathcal{T}_{TF}[\rho] := \int d\mathbf{r} C_F \rho^{5/3}(\mathbf{r}) \quad (4.2)$$

$$\mathcal{J}_{TF}[\rho] := \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} \quad (4.3)$$

incluye además a la *energía de Hartree* ($\mathcal{J}_{TF}[\rho]$) que puede interpretarse como la interacción de la densidad de carga consigo misma. Esto no da buenos resultados. Conviene expresar a la densidad de un sistema de dos partículas como

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \rho(\mathbf{r}_1)\rho(\mathbf{r}_2) + \rho(\mathbf{r}_1)h_{xc}(\mathbf{r}_1; \mathbf{r}_2)$$

Integrando a ambos miembros sobre todo el espacio usando a \mathbf{r}_2 como variable de integración

$$(N-1)\rho(\mathbf{r}_1) = N\rho(\mathbf{r}_1) + \rho(\mathbf{r}_1) \int d\mathbf{r}_2 h_{xc}(\mathbf{r}_1; \mathbf{r}_2)$$

Entonces

$$\int d\mathbf{r}_2 h_{xc}(\mathbf{r}_1; \mathbf{r}_2) = -1$$

A la función $h_{xc}(\mathbf{r}_1; \mathbf{r}_2)$ se la llama *hueco de intercambio-correlación*. En el caso particular en que las partículas no interactúen entre sí, se le llama *hueco de intercambio*.

La contemplación del intercambio condujo a modificar la funcional de Thomas-Fermi. Se lo hizo sumando la funcional de intercambio de Dirac, cuya expresión es

$$\mathcal{K}_D[\rho] := \frac{3}{4} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} \int d\mathbf{r} \rho^{\frac{4}{3}}(\mathbf{r})$$

Estas funcionales proveen una mala descripción de la energía cinética en sistemas no homogéneos. En caso de intentar aplicarlos a moléculas, no predicen los enlaces.

4.3. Teoría de Kohn-Sham

La funcional exacta es desconocida. Con la teoría de Kohn-Sham[20] se utiliza un sistema hipotético como referencia, que representa a la mayor parte de la energía del sistema. Así sólo es necesario realizar aproximaciones en términos correctivos.

El sistema de referencia utilizado es un sistema de partículas que no interactúan entre sí. En cuyo caso, la funcional de Levy-Lieb es

$$\mathcal{F}_{LL}[\rho] = \langle \Psi_\rho | \mathcal{T} + \mathcal{W} | \Psi_\rho \rangle = \langle \Psi_\rho | \mathcal{T} | \Psi_\rho \rangle = \mathcal{T}_s[\rho]$$

donde se definió a la funcional de energía cinética de Kohn-Sham $\mathcal{T}_s[\rho]$. Esta puede ser evaluada en forma exacta, dando mejores resultados que \mathcal{F}_{TF} .

En el marco de esta teoría la funcional exacta es separada según la descomposición de Kohn-Sham,

$$\mathcal{F}[\rho] = \mathcal{T}[\rho] + \mathcal{W}[\rho] = \mathcal{T}_s[\rho] + \mathcal{J}[\rho] + \mathcal{E}_{ex}[\rho]$$

Al último término se lo conoce como funcional de intercambio y correlación. La energía corresponderá a

$$E[\mathcal{V}] = \min_{\rho} (\mathcal{T}_s[\rho] + \mathcal{J}[\rho] + \mathcal{E}_{ex}[\rho] + \langle \mathcal{V} | \rho \rangle)$$

Dado que se consideró que las partículas no interactúan entre sí, el estado fundamental del sistema puede expresarse en términos de un determinante de Slater de funciones monoelectrónicas ϕ_i , por lo que

$$\begin{aligned} \left[-\frac{1}{2}\nabla^2 + \mathcal{V}_s(\mathbf{r})\right]\phi_i(\mathbf{r}) &= \epsilon_i\phi_i(\mathbf{r}) \\ \sum_i^N |\phi_i(\mathbf{r})|^2 &= \rho(\mathbf{r}) \end{aligned}$$

El potencial $\mathcal{V}_s(\mathbf{r})$ representa a un potencial externo, que da origen a la misma densidad electrónica que el sistema real, y a partir de lo expuesto en esta sección puede demostrarse que

$$\begin{aligned} \mathcal{V}_s(\mathbf{r}) &= \mathcal{V}(\mathbf{r}) + \mathcal{V}_J(\mathbf{r}) + \mathcal{V}_{xc}(\mathbf{r}) \\ \mathcal{V}_J(\mathbf{r}) &:= \frac{\delta \mathcal{J}[\rho]}{\delta \rho(\mathbf{r})} = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\ \mathcal{V}_{xc}(\mathbf{r}) &:= \frac{\delta \mathcal{E}_{xc}[\rho]}{\delta \rho(\mathbf{r})} \end{aligned}$$

En principio las ecuaciones de Kohn-Sham proveen la densidad exacta. Los orbitales obtenidos minimizan la energía cinética del sistema de partículas que no interactúan entre sí. La función de onda corresponde a un determinante de Slater formado a partir de estos orbitales.

Para realizar el procedimiento es necesario tener la forma explícita de $\mathcal{E}_{xc}[\rho]$. Existen muchas aproximaciones para esto. Es costumbre dividirlo en dos términos

$$\mathcal{E}_{xc}[\rho] = \mathcal{E}_x[\rho] + \mathcal{E}_c[\rho]$$

Al primer término se lo denomina funcional de intercambio, y al segundo funcional de correlación. El aporte del primero a la energía es mucho mayor que el del segundo. La posibilidad de separación en sumandos no tiene una justificación teórica, pero ha mostrado ser conveniente en la práctica.

4.4. B3LYP

Para realizar cálculos es necesario utilizar alguna forma explícita para \mathcal{E}_{xc} . Previamente se comentaron formas explícitas para sistemas homogéneos. La densidad y los potenciales

no son homogéneos en las moléculas. Una aproximación que permite utilizar dichas formas para cálculos de sistemas moleculares es la *aproximación de la densidad local* (LDA). Esta consiste en considerar a los sistemas inhomogéneos como un conjunto de pequeños sistemas homogéneos contiguos, en los cuales la densidad y el potencial son constantes. Se aplican los resultados para sistemas homogéneos a cada subsistema. Se busca definir a los subsistemas de modo que sean infinitamente pequeños. Como funcional de intercambio se usa la de Dirac. Como funcional de correlación se utiliza alguna que sólo dependa de la densidad, como la de Vosko–Wilk–Nusair (VWN)[21] o la de Perdew-Zunger (PZ81)[22].

El avance natural por sobre el modelo LDA es la inclusión de la dependencia de la funcional con el gradiente de la densidad. Esto se aplica a la funcional de intercambio y correlación. En las aproximaciones del gradiente generalizadas GGA, la funcional toma la forma

$$\mathcal{E}_{xc}^{\text{GGA}}[\rho] = \int d\mathbf{r} f(\rho_\alpha, \rho_\beta, \xi_{\alpha\alpha}, \xi_{\beta\beta}, \xi_{\alpha\beta})$$

$$\xi_{ab} = \nabla \rho_a \cdot \nabla \rho_b$$

El diseño de estas funcionales es guiado mediante restricciones, como satisfacer ciertas condiciones de contorno, alguna relación matemática conocida, y ajustes a valores teóricos y o experimentales. Suelen incluir algún conjunto de parámetros.

Becke propuso la funcional de intercambio[23]

$$E_x[\rho] = \sum_\sigma \int d\mathbf{r} \rho_\sigma (\mathcal{E}_{x\sigma}^{\text{LDA}} + \Delta\mathcal{E}_{x\sigma}^{\text{B88}})$$

$$\Delta\mathcal{E}_{x\sigma}^{\text{B88}} := -\beta \rho_\sigma^{1/3} \frac{s_\sigma^2}{1 + 6\beta s_\sigma \sinh^{-1} s_\sigma}$$

$$s_\sigma := \frac{|\nabla \rho_\sigma|}{\rho_\sigma^{4/3}}$$

El parámetro β es empírico y su valor (0,0042) fue obtenido por ajuste al intercambio de Hartree-Fock de gases nobles.

Para la energía de correlación Lee, Yang y Parr propusieron[24]²

²Esta forma funcional es válida para sistemas de capa cerrada.

$$\begin{aligned}\mathcal{E}_c[\rho]^{\text{LYP}} &= -a(1 + d\rho^{-1/3})^{-1} - ab\omega\rho[C_F\rho^{8/3} + |\nabla\rho|^2(5/12 - 7\delta/72) - (11/24)\rho^2|\nabla\rho|^2] \\ \omega &:= \frac{\exp(-c\rho^{-1/3})}{1 + d\rho^{-1/3}}\rho^{-11/3} \\ \delta &:= c\rho^{-1/3} + \frac{d\rho^{-1/3}}{1 + d\rho^{-1/3}}\end{aligned}$$

Esta funcional tiene 4 parámetros ajustables (a , b , c , d). Suele usarse en combinación con la corrección B88 al intercambio³.

Es posible usar el intercambio exacto ($\mathcal{E}_x^{\text{HF}}$) en vez de la funcional de intercambio. Esto no suele dar buenos resultados porque las funcionales de correlación usualmente no describen bien la correlación estática. Una técnica que ha dado buenos resultados es mezclar una funcional de intercambio con el intercambio exacto. Las funcionales creadas de esta manera reciben el nombre de funcionales híbridas. Nótese que el método se torna semiempírico, con la excepción de PBE0[25, 26]. Por mucho, la más utilizada para cálculos de sistemas moleculares es la funcional híbrida B3LYP[27, 28], cuya forma es

$$\mathcal{E}_{xc}^{\text{B3LYP}} = (1 - a)\mathcal{E}_{\text{Dirac}} + a\mathcal{E}_x^{\text{HF}} + b\mathcal{E}_x^{\text{B88}} + (1 - c)\mathcal{E}_c^{\text{VWN}} + c\mathcal{E}_c^{\text{LYP}}$$

donde la funcional de correlación $\mathcal{E}_c^{\text{VWN}}$ (de Vosko–Wilk–Nusair (VWN)) es una funcional de tipo LDA basada en simulaciones, y $(a;b;c)=(0,20;0,71;0,81)$.

4.5. Cálculo de propiedades magnéticas con DFT

Las constantes de apantallamiento y de acoplamiento indirecto pueden ser obtenidas mediante DFT. Existen distintas expresiones dependiendo del formalismo utilizado. Sychrovský, Gräfenstein y Cremer obtuvieron[29] las siguientes expresiones para las constantes de acoplamiento indirecto

³Esto da origen a la funcional BLYP

$$\begin{aligned}
\mathbf{K}_{KL}^{dso} &= \sum_i^{\text{occ}} \langle \phi_i | \mathbf{D}_{KL} + h_{KL}^{dso} | \phi_i \rangle \\
\mathbf{K}_{KL}^{psd} &= -\frac{2}{3} \sum_i^{\text{ocu}} \sum_a^{\text{virt}} (\epsilon_a - \epsilon_i)^{-1} \langle \phi_i | \mathbf{h}_K^{psd} | \phi_i \rangle \cdot \langle \phi_i | \mathbf{h}_L^{psd} + \mathbf{F}_L^{psd} | \phi_i \rangle \\
\mathbf{K}_{KL}^{fc} &= -\frac{2}{3} \sum_i^{\text{ocu}} \sum_a^{\text{virt}} (\epsilon_a - \epsilon_i)^{-1} \langle \phi_i | \mathbf{h}_K^{fc} | \phi_i \rangle \cdot \langle \phi_i | \mathbf{h}_L^{fc} + \mathbf{F}_L^{fc} | \phi_i \rangle \\
\mathbf{K}_{KL}^{sd} &= -\frac{2}{3} \sum_i^{\text{ocu}} \sum_a^{\text{virt}} (\epsilon_a - \epsilon_i)^{-1} \langle \phi_i | \mathbf{h}_K^{sd} | \phi_i \rangle \cdot \langle \phi_i | \mathbf{h}_L^{sd} + \mathbf{F}_L^{sd} | \phi_i \rangle
\end{aligned}$$

donde los operadores \mathbf{F}_K^X deben ser obtenidos en forma auto-consistente a partir de

$$\begin{aligned}
\mathbf{F}_K^X &= \sum_i^{\text{occ}} \int d^3\mathbf{r} \frac{\delta F}{\delta \phi_i(\mathbf{r})} \phi^{(K),X}(\mathbf{r}) \\
|\phi^{(K),X}(\mathbf{r})\rangle &= - \sum_a^{\text{virt}} \frac{\langle \phi_i | \mathbf{h}_L^X + \mathbf{F}_K^X | \phi_i \rangle}{\epsilon_a - \epsilon_i} |\phi_i(\mathbf{r})\rangle
\end{aligned}$$

Estas guardan mucha semejanza con las de Ramsey. Nótese sin embargo que los valores esperados son formados con los orbitales de Kohn-Sham. También que las energías aquí se refieren a las energías de estos orbitales. En cuanto a aspectos prácticos, es importante notar que se obtienen mejores resultados con funcionales híbridos que con funcionales puros.

4.5.1. Viabilidad y conveniencia

Los cálculos basados en DFT son, normalmente, más precisos que aquellos basados en Hartree-Fock (HF). Si bien los últimos son útiles para apantallamientos, sus resultados no lo son para la predicción de constantes de acoplamiento indirecto. Sin embargo, mientras que Hartree-Fock provee una aproximación basada en primeros principios, la justificación teórica del uso de DFT para estas propiedades es más débil. Aún así, la experiencia muestra que se obtienen resultados aceptables. En un estudio de casi 30 moléculas orgánicas[30], utilizando la base 6-311+G(2d,p), se encontró que para ^{13}C el error absoluto medio (MAE) de los corrimientos químicos fue de 3,6 para B3LYP y 9,2 para HF. Para ^{15}N los respectivos MAE fueron 16,9 y 35,3⁽⁴⁾.

⁽⁴⁾Los corrimientos son relativos al trimetilsilano (TMS) para ^{13}C y al amoníaco para ^{15}N .

4.6. Teoría del funcional de la densidad y la corriente

Dado un sistema bajo el efecto de un campo magnético, el hamiltoniano que lo representa incluye, como se discutió extensivamente, al potencial vector magnético. En el caso de un electrón despojado de su espín, el hamiltoniano toma la forma

$$H(\mathcal{V}, \mathbf{A}) = H_0 + \mathcal{V} + \frac{1}{2}\mathbf{A}^2 + \mathbf{A} \cdot \mathbf{p}$$

El término no lineal en \mathbf{A} puede generar la no concavidad de la energía respecto a \mathbf{A} . Otra consecuencia importante es la pérdida de linealidad del hamiltoniano. Se ha mostrado que lo mismo se cumple para sistemas de capa cerrada más complejos [31].

Ambos problemas pueden evitarse realizando una reparametrización con la variable $\mathbf{u} = \mathcal{V} + \frac{1}{2}\mathbf{A}^2$ [32],

$$H(\mathcal{V}, \mathbf{u}) = H_0 + \mathbf{u} + \mathbf{A} \cdot \mathbf{p}$$

Haciendo esto, es posible proceder en forma análoga a lo realizado para DFT, y obtener los correspondientes principios variacionales de Lieb y Hohenberg–Kohn

$$\begin{aligned} E[\mathbf{u}, \mathbf{A}] &= \inf_{\rho, \kappa} (\mathcal{F}[\rho, \kappa] + \langle \mathbf{u} | \rho \rangle + \langle \mathbf{A} | \kappa \rangle) \\ \mathcal{F}[\rho, \kappa] &= \sup_{\mathbf{u}, \mathbf{A}} (E[\mathbf{u}, \mathbf{A}] - \langle \mathbf{u} | \rho \rangle - \langle \mathbf{A} | \kappa \rangle) \end{aligned}$$

Nótese que la funcional universal depende también de la *densidad de corriente paramagnética* κ ,

$$\kappa := \Re(\Psi(\mathbf{r})\mathbf{p}\Psi(\mathbf{r}))$$

Al igual que la densidad electrónica, la densidad de corriente paramagnética queda determinada por la estructura electrónica del sistema⁵. Esto no es así para la densidad de corriente \mathbf{j} , la cual es un observable físico. κ surge como variable natural por conjugación convexa. Por tal motivo muchas veces se la incluye, en lugar de \mathbf{j} , en este formalismo al cual se lo llama *Teoría de la funcional de la densidad y la corriente* (CDFT).

⁵Porque κ queda determinada por la función de onda, y esta a su vez por la estructura electrónica.

Este formalismo es muy reciente, hubo mucho desarrollo en los últimos años. En 2012 Tellgren *et al.* propusieron[32] que no eran correctos argumentos previos que pretendían demostrar la existencia del teorema de Hohenberg–Kohn para CDFT mediante el uso de la densidad corriente. En 2014 Laestadius *et al.* hicieron una demostración similar. Además mostraron que si dos sistemas tienen el mismo par (ρ, \mathbf{j}) \mathcal{V} -representable, los sistemas comparten el mismo conjunto de estados fundamentales degenerados. También remarcaron que para estas dos variables, no hay demostraciones a favor o en contra de un principio de HK para sistemas con más de una partícula[33]. En 2015 Laestadius *et al.* mostraron que la funcional generalizada de Hohenberg-Kohn⁶ puede ser mínima para densidades que no corresponden al estado fundamental de los correspondientes potenciales. A fines de 2015 Helgaker *et al.* [34] lograron establecer las condiciones que permiten que todas las densidades físicas sean recuperadas por las funcionales, y que no se generen densidades no físicas. Recientemente, en agosto de 2016 Sahni *et al.* demostraron[35] un principio de HK para (ρ, \mathbf{j}) . En este caso, el par (ρ, \mathbf{j}) determina los potenciales \mathcal{V} y \mathbf{A} , a menos de una constante y el gradiente de una función escalar. Este principio es válido para sistemas plurieléctricos pero está sujeto a restricciones de conservación de los momentos angulares.

Desde un punto de vista teórico, la inclusión de alguna densidad de corriente es necesaria. Desde un punto de vista práctico, esto debe juzgarse a partir de la calidad de los resultados obtenidos. Esto ha sido analizado recientemente (2015) en términos de la magnitud de la contribución de la densidad de corriente a los apantallamientos[36]. Por comparación con cálculos *ab initio* de alto nivel los autores encontraron una mejora, tras la inclusión de la densidad de corriente, de aproximadamente 3 a 4 ppm en los MAE, y una disminución de 2 ppm en la desviación estándar⁷.

No se han realizado cálculos CDFT para esta tesis. Tenemos acceso a programas informáticos capaces de realizar los cálculos con DFT, y los últimos resultan suficientemente precisos. El programa London[38], capaz de realizar cálculos CDFT, aún no es accesible a la comunidad.

⁶Que incluye a la densidad de corriente.

⁷Esto no se observó para la funcional KT2[37], la cual fue diseñada para la predicción de estas propiedades.

4.7. Ideas principales sobre el capítulo

Se realizó una introducción a la teoría del funcional de la densidad, porque en ella se basaron los cálculos de propiedades magnéticas de esta tesis. El teorema de Hohenberg y Kohn demuestra que, para estados fundamentales de sistemas N electrónicos no sujetos a potenciales vectoriales, a cada densidad electrónica le corresponde sólo un⁸ potencial externo. Puede definirse entonces una funcional de la densidad electrónica (funcional de HK), que según el principio variacional, permitiría por minimización obtener la energía del sistema. Surgen dos problemas, el primero es la caracterización del dominio de funciones de esta funcional⁹, el otro es el desconocimiento de cual es la forma de esta funcional. El primero, fue esencialmente resuelto mediante la funcional de Lieb, lo cual habilita la utilización de técnicas estándar de optimización de funciones. El segundo problema no ha sido resuelto. Por lo tanto, la teoría es en la práctica semi-empírica. Se comentaron algunas de las primeras aproximaciones a sistemas sencillos. Buenas aproximaciones se pueden realizar en el marco de la teoría de Kohn-Sham, en la que se utiliza como referencia a un sistema de electrones no interactuantes, cuyas propiedades pueden ser evaluadas con facilidad. Para ello se incorporan orbitales a la teoría. La funcional es dividida en términos, en los que sólo un pequeño aporte es desconocido. Este es asignado a los funcionales de intercambio y correlación, cuyas formas exactas no se conocen. Se comentó sobre el funcional híbrido (en tanto que una parte de su intercambio corresponde al intercambio exacto) B3LYP, que es el usado para los cálculos en esta tesis. Se mostraron expresiones válidas para el cálculo de las constantes de acoplamiento en el marco de DFT. Se comentó que, cuando se incluye un potencial vector en el hamiltoniano, se vuelve teóricamente necesario el agregado de alguna densidad de corriente como argumento de la funcional. Esto da origen a la teoría CDFT, sobre la cual ya se comentó someramente respecto a los avances recientes.

⁸Salvo constante aditiva

⁹Las ecuaciones serían válidas sólo para densidades correspondientes a estados fundamentales, y es muy difícil saber *a priori* cuando una densidad pertenece a este espacio.

Parte III

TEORÍA QSPR

Capítulo 5

QSPR

5.1. Introducción

Se llama **modelos QSPR** (del inglés **Quantitative Structure Property Relationships**) a modelos estadísticos de regresión o clasificación. Sus variables predictoras son valores numéricos obtenidos como imágenes de funciones de variables asociadas a propiedades fisicoquímicas de especies químicas, o de algoritmos aplicados sobre representaciones formales de las estructuras de las mismas. Sus variables de respuesta están asociadas a propiedades fisicoquímicas o valores categóricos.

Definición 5.1.1. Se llama **descriptor** a cualquier número real¹ asociado con una especie química y obtenido mediante un proceso estandarizado con el objetivo de ser usado en el modelo QSPR como un predictor potencial. También se llama así al algoritmo con el cual se genera unívocamente dicho número a partir de cálculos teóricos, o de una representación de la estructura y o información experimental.

Un modelo QSPR se construye estableciendo una relación entre las variables independientes (los descriptores), con las variables dependientes (magnitud o categoría que interese predecir), empleando información correspondiente a un conjunto de especies químicas de las cuales ambos conjuntos de variables son conocidos.

Una vez construido el modelo, se lo puede utilizar con el objeto de estimar el valor numérico o la categoría correspondiente a especies químicas no presentes en el conjunto utilizado para su construcción.

¹El requerimiento de pertenencia al conjunto de los números reales se debe a que estos son usados para la predicción de propiedades químicas o biológicas, normalmente a partir de una combinación lineal de los mismos.

La racionalidad de estos modelos se basa en la hipótesis que consiste en considerar que dichas relaciones existen. Para que sean útiles en la práctica debe existir información suficiente para construirlos con una precisión aceptable. También debe ocurrir que el cálculo de los descriptores usados en el modelo sea más conveniente que la obtención por otros métodos de aquello que interesa predecir.

Comentario 5.1.2. La creación de un *modelo estadístico* conlleva suposiciones que son utilizadas para la generación de datos potencialmente observables. A fines prácticos puede entenderse a estos modelos como representaciones formales de procesos con los cuales se crean dichos datos. Una característica deseable de los primeros es que sean suficientemente sencillos como para permitir una idea intuitiva de la entidad a la cual, con los datos involucrados, se describe.

Formalmente puede representárselos como una dupla (Ω, \mathcal{P}) , en la que Ω es el *espacio muestral* (el conjunto compuesto por todos los resultados obtenibles) y \mathcal{P} es un conjunto de distribuciones de probabilidad sobre Ω . Se requiere que exista al menos una distribución en \mathcal{P} que sea una buena aproximación a una distribución verdadera, de la cual se asume su existencia. El conjunto \mathcal{P} contiene distribuciones creadas a partir de un conjunto de parámetros Θ , $(\mathcal{P} = \{p_\theta : \theta \in \Theta\})$ [39].

La naturaleza de estos métodos es más general, siendo aplicables en cualquier campo en el que se requiera determinar un conjunto de variables para obtener un **estimador** o un **predictor**.

Los conceptos de estimador y predictor no suelen diferenciarse por estar íntimamente relacionados. A lo largo de este texto no se hará tal distinción, pero es de interés establecer una clara diferencia.

La información disponible para la construcción de un modelo es normalmente pares (\mathbf{x}, \mathbf{y}) . El espacio muestral Ω constituye el conjunto de todos los pares (\mathbf{x}, \mathbf{y}) posibles. Luego deben realizarse suposiciones, por ejemplo, que se puede usar una relación lineal para aproximar a las variables,

$$\mathbf{y} = \beta \mathbf{x} + \alpha + \epsilon$$

En esta expresión α y β representan las matrices que contienen los coeficientes de la relación lineal. Nótese que todos los elementos en α (α) y en β (β) son iguales. ϵ representa la matriz de los errores. Otra suposición realizada puede ser que los errores son independientes e idénticamente distribuidos en forma Gaussiana.

Se tienen entonces conjuntos de parámetros $\theta_i = (\alpha_i, \beta_i, \sigma_i)$ siendo sigma una medida de la dispersión. Por cada conjunto θ_i existe una distribución p_{θ_i} de los errores. Luego, si $\Theta = \{\theta_i\}$, entonces $\mathcal{P} = \{p_\theta : \theta \in \Theta\}$.

Se supone que existen valores verdaderos (pero desconocidos) de α y β . Verdaderos en tanto que son idóneos en algún sentido, en particular generando errores pequeños.

Un estimador es un procedimiento matemático que asigna, a cada posible conjunto de observaciones un $\hat{\theta}$. Un par $(\hat{\alpha}, \hat{\beta})$ en el ejemplo (σ permanece desconocido y debe ser estimado). Nótese que el estimador es una función y no es una variable aleatoria. Sin embargo, las estimaciones son variables aleatorias dado que se construyen utilizando un estimador junto a variables aleatorias.

En cambio, un predictor de las variables es una función $p(x)$ que asigna un valor tentativo de otra variable aleatoria, relacionada con el conjunto. Por ejemplo, un predictor para la variable y podría ser $\hat{\beta}x + \hat{\alpha}$.

Esencialmente modelos QSPR son modelos de regresión o clasificación donde la naturaleza de las variables a relacionar es la recién descrita. Por lo tanto una introducción a QSPR consiste en una introducción a los aspectos de regresión y clasificación más utilizados en este campo.

Estos modelos son utilizados ampliamente en ámbito de las ciencias biológicas para la predicción de actividad biológica, en cuyo caso dichos modelos son llamados **modelos QSAR** (del inglés **Quantitative Structure-Activity Relationship**).

Comentario 5.1.3. Existe cierta controversia acerca de la existencia o no de causalidad en las relaciones reportadas de estudios QSAR/QSPR, ver por ejemplo[40, 41]. Esto amerita algunos comentarios.

La *causalidad* es una propiedad de la relación entre los elementos de un conjunto de eventos llamados *causas*, y los de otro conjunto de eventos llamados *efectos*.

Se acepta que las causas originan, y por lo tanto preceden, a sus efectos. Esto implica un orden temporal. De acuerdo con la teoría de la relatividad, la simultaneidad de los eventos es dependiente del sistema de referencia. Entonces, el orden temporal puede ser dependiente del sistema de referencia. Para poder clasificar a eventos como causas y efectos se impone el requerimiento de que el orden temporal sea el mismo en todos los sistemas de referencia inerciales.

Una cantidad invariante ante el sistema de referencia es el llamado *intervalo*, $I = (c\Delta t)^2 - (\Delta x^2 + \Delta y^2 + \Delta z^2)$, referido a los estados de tiempo y espacio de dos eventos. Se llama intervalo espacial si $I < 0$, y temporal si $I > 0$. Si el intervalo entre dos eventos \mathcal{A} y \mathcal{B} es temporal, y en algún sistema de referencia inercial \mathcal{A} es anterior a \mathcal{B} , entonces en todos los sistemas inerciales \mathcal{A} es anterior a \mathcal{B} . Esto hace en principio posible la existencia de una relación causal entre \mathcal{A} y \mathcal{B} . En caso que el intervalo sea espacial, el orden temporal entre los eventos \mathcal{A} y \mathcal{B} es dependiente del sistema inercial. En cuyo caso, no existe una relación causal posible.

No es sencillo determinar la causalidad de una relación ni la entidad del agente causante. La existencia y conocimiento de dicha relación son normalmente anhelados. Esto puede ocurrir debido a que suele creerse que así se reduce la probabilidad de obtener correlaciones fortuitas. Se debe tener presente que existen métodos para analizar si una relación es fortuita o no lo es. En casos sencillos la determinación, tanto de la relación causal como de la correlación, se basa en la experiencia.

Una relación de causalidad no implica una buena correlación. Es posible que se conozca la existencia de una relación causal, pero que se desconozca cómo cuantificar el efecto a partir de la causa. La existencia de otros efectos no contemplados puede dar origen a una mala correlación. En relaciones bien determinadas cuantitativamente no siempre es posible atribuir una relación de causa efecto. Conocer dicha relación no es necesariamente útil para obtener un modelo predictivo.

En QSPR se relacionan propiedades de especies químicas. Bajo este nivel de observación, dichas propiedades son distintas manifestaciones de la naturaleza de la especie en cuestión. Por lo tanto no son eventos sobre los cuales el concepto de relación causa-efecto, discutido arriba, sea aplicable. Dado que ambas propiedades son consecuencia de la estructura a un nivel inferior, la existencia de algún tipo de correlación no fortuita es plausible (pero no necesaria). En esencia, esto es lo que sustenta el uso de QSPR. Naturalmente, debe procurarse que la correlación encontrada sea relevante estadísticamente.

5.2. Descriptores

Ya se definió a los descriptores con la definición 5.1.1. En ella se hizo referencia a representaciones de las estructuras de las especies químicas. Aunque la bibliografía no es explícita al respecto, se considera implícitamente que la estructura de la especie química es aquella correspondiente a la descripción de esta obtenida mediante la utilización de la aproximación de Born-Oppenheimer². De estas representaciones se requiere que sean obtenidas a partir de reglas bien establecidas. Es costumbre en la bibliografía clasificar a los descriptores de acuerdo a la naturaleza de la representación usada para obtenerlos, por ello se describirá aquí dicha clasificación.

Una representación sencilla de una especie química está dada por su fórmula química. A aquellos descriptores obtenidos sin utilizar más información estructural que la provista por estas representaciones se los clasifica como *descriptores 0D*. Ejemplos de esta clase de descriptores son el número de átomos de oxígeno de una especie química y la masa de la misma.

Otra representación consiste en una lista de representaciones de fragmentos de la especie química. Estos fragmentos son normalmente grupos funcionales o sustituyentes que pudieran resultar de interés. A los descriptores obtenidos utilizando como única información estructural aquella deducible de estas representaciones se los llama *descriptores 1D*. Un ejemplo de estos es el número de grupos carboxilo de una especie química.

La estructura de una especie química puede ser representada por un grafo. De estos se puede deducir la conectividad atómica incluida la especificación del orden de enlace. En la mayoría de los casos la dimensión topológica de estos grafos es 2, por lo que a los descriptores contruidos con información estructural que sólo proviene de estos grafos se les llama *descriptores 2D*.

En el caso que la representación de la especie química usada para la obtención de la información estructural empleada en la construcción de un dado descriptor, consista en una lista o conjunto de las posiciones de cada núcleo que la compone, se dice del descriptor que es un *descriptor 3D*. En este caso, se puede argumentar en favor de la pertinencia del nombre remarcando que, para contemplar todos los casos posibles se requiere un espacio de dimensión igual o mayor que 3 para representar la posición de un núcleo cualquiera.

²Nótese que esto es distinto a decir que la representación estructural a partir de la cual se crean los descriptores es la obtenida mediante la aproximación de Born-Oppenheimer. Se pueden emplear representaciones más sencillas, tales como la fórmula química.

5.2.1. Descriptores base

En esta sección se describe a los aspectos generales de los descriptores que fueron utilizados en esta tesis como base para la creación de descriptores dependientes de propiedades magnéticas.

3D-MoRSE

Su nombre proviene la frase en inglés *Molecule Representation of Structures based on Electron diffraction*, dado que su expresión matemática surgió de modificar y simplificar una función matemática útil en el estudio de difracción de electrones. La expresión general para esta familia de descriptores es:

$$I(s) = \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}}$$

Donde N representa al número de átomos de la molécula, r_{ij} a la distancia euclídea entre los átomos i y j y A_k al valor de una propiedad atómica del átomo k . La función $I(s)$ corresponde a distintos descriptores dependiendo del valor de s elegido. Se la evalúa para s natural en el rango $\{s \in \mathbb{N} | 1 \leq s \leq 31\}$, y para $s \rightarrow 0$.

RDF

Su nombre es la abreviatura de *Radial Distribution Function*. Estos descriptores corresponden a las imágenes de la función $g : \mathbb{R} \rightarrow \mathbb{R}$, definida por

$$g(R) := f \sum_{i=2}^N \sum_{j=1}^{i-1} A_i A_j e^{-\beta(R-r_{ij})^2}$$

En la que $f \in \mathbb{R}$ representa a un factor de escala, $\beta \in \mathbb{R}^+$ a un factor de amortiguación. Las definiciones del resto de los símbolos son las usadas para los descriptores 3D-MoRSE.

GETAWAY

Su nombre hace referencia a *GEometry, Topology, and Atom-Weights Assembly*. Estos descriptores son obtenidos a partir de dos matrices. Una de ellas, \mathbf{H} , llamada *Molecular Influence Matrix*, con elementos h_{ij} definida como:

$$\mathbf{H} = \mathbf{M} \cdot (\mathbf{M}^T \cdot \mathbf{M})^{-1} \cdot \mathbf{M}^T$$

donde \mathbf{M} representa la matriz (de $N \times 3$) de las coordenadas x, y, z de los átomos.

La otra matriz, \mathbf{R} , es una matriz cuadrada definida a partir de sus elementos R_{ij} como

$$R_{ij} = \begin{cases} \frac{\sqrt{h_{ii}h_{jj}}}{r_{ij}} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

El elemento $0 \leq h_{ii} \leq 1$ correspondientes al átomo i , tiende a tener valores mayores a medida que el átomo se aleja del centro de la molécula. Los elementos h_{ij} tienden a ser mayores a medida que los átomos i y j se hallan más alejados entre sí.

Existen varios descriptores de esta clase, que permiten la inclusión de valores de propiedades atómicas. Se denotará por $\delta(i, j)$ a la delta de Kronecker. Se define aquí a la función $\delta(i, j, k) : \mathbb{N} \times \mathbb{N} \times \mathbb{R} \rightarrow \{0, 1\}$ con

$$\delta(i, j, k) = \begin{cases} 1 & \text{si } i = j \text{ y } k > 0 \\ 0 & \text{si } i \neq j \text{ o } k \leq 0 \end{cases}$$

Denotando por d_{ij} al menor número de enlaces covalentes a través de los cuales puede conectarse al átomo i con el j . Y d al máximo d_{ij} de la molécula. Las definiciones de aquellos descriptores de esta clase que serán utilizados en esta tesis se escriben a continuación

$$\text{HATS}_k(A) := \sum_{i=1}^{N-1} \sum_{j \geq i}^N A_i A_j h_{ii} h_{jj} \delta(k, d_{ij}) \quad (5.1)$$

$$\text{HATS}(A) := \text{HATS}_0(A) + 2 \sum_{k=1}^d \text{HATS}_k(A) \quad (5.2)$$

$$\text{H}_k(A) := \sum_{i=1}^{N-1} \sum_{j \geq i}^N A_i A_j h_{ij} \delta(k, d_{ij}, h_{ij}) \quad (5.3)$$

$$\text{HT}(A) := \text{H}_0(A) + 2 \sum_{k=1}^d \text{H}_k(A) \quad (5.4)$$

$$\text{R}_k(A) := \sum_{i=1}^{N-1} \sum_{j \geq i}^N A_i A_j R_{ij} \delta(k, d_{ij}) \quad (5.5)$$

$$\text{RT}(A) := 2 \sum_{i=1}^{N-1} \sum_{k=1}^d \text{R}_k(A) \quad (5.6)$$

WHIM

Se llama WHIM, del inglés *Weighted Holistic Invariant Molecular descriptors*, a una clase de descriptores relacionados entre sí a través del origen de sus algoritmos, dado por la obtención de una matriz de covarianza pesada S , definida a partir de sus elementos $S_{p,q}$, definidos por:

$$S_{p,q} := \frac{\sum_{i=1}^N A_i (p_i - \bar{p})(q_i - \bar{q})}{\sum_{i=1}^N A_i}$$

Donde A_i es el valor de una magnitud A del átomo i . p y q representan a los valores de las coordenadas atómicas de posición x , y o z , y \bar{p} y \bar{q} a los valores medios de las correspondientes coordenadas.

Entre los muchos descriptores de esta clase que han sido creados, se escogió la forma de un subconjunto de ellos. El criterio de elección fue la simplicidad de implementación e interpretación. Siendo $\{\lambda_i\}$ los autovalores de la matriz S , se utilizarán los descriptores

$$\lambda_i \quad (i = 1, 2, 3) \quad (5.7)$$

$$T \quad := \quad \lambda_1 + \lambda_2 + \lambda_3 \quad (5.8)$$

$$A \quad := \quad \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_2 \lambda_3 \quad (5.9)$$

$$\theta_i \quad := \quad \frac{\lambda_i}{T} \quad (5.10)$$

ATS

Se llama ATS, del inglés *Autocorrelation of a Topological Structure*, a los descriptores definidos por

$$ATS_k(A) = \frac{1}{2} \sum_{i=1}^N \sum_{j \geq i}^N A_i A_j \delta(k, d_{ij})$$

En este trabajo se utilizaron valores enteros de k entre 0 y 3.

5.3. Construcción de modelos

Para la construcción de un modelo QSPR es necesario disponer de al menos un conjunto de especies químicas de las cuales se conozca al menos un tipo de representación

de sus estructuras y la propiedad a predecir. A un conjunto con estas características que sea utilizado para obtener valores numéricos probables de los coeficientes del modelo, se le llama **conjunto de ajuste**.

A partir de dichas representaciones se generan los descriptores correspondientes a las especies químicas del conjunto de ajuste. La generación se hace comúnmente con algún programa informático. Existen varios programas informáticos capaces de calcular descriptores a partir de una lista de archivos en los que se encuentre codificada la información de la representación de las especies químicas, usando formulas matemáticas como las vista en la sección anterior. Los formatos de archivo admitidos no difieren por lo común de aquellos usados habitualmente en química computacional, tales como Molfile, Protein Data Bank (pdb) o XYZ. Desde el punto de vista del usuario, las principales diferencias entre los programas informáticos disponibles deberían (según el criterio del autor) radicar en la cantidad, cualidades y calidad de descriptores que son capaces de generar, y en el costo y la licencia de su uso.

Los descriptores generados son almacenados en una matriz de $m \times d$ (o su transpuesta según conveniencia), donde el m representa al número de especies químicas del conjunto de ajuste, y d al número de descriptores diferentes generados para cada especie química. Cada elemento correspondiente a la fila i y columna j representa al descriptor número j de la especie química i .

Cada valor de la propiedad de interés correspondiente a las especies químicas del conjunto es guardado en una matriz de $m \times 1$ (o $1 \times m$).

Al menos varias decenas de descriptores por molécula son usados normalmente en los estudios de QSPR. Actualmente suelen emplearse unos cinco mil.

El número de especies químicas utilizado debe ser suficientemente grande para que el modelo sea estadísticamente confiable, es esperable al menos que sea del orden de las decenas.

La relación más utilizada en modelos QSPR entre descriptores y la propiedad que interese predecir se expresa como una igualdad entre el valor de dicha predicción y una combinación lineal de descriptores.

Representando y_i al valor predicho para la magnitud y asociada a la especie química i , $\{d_{ij}\}$ a un conjunto de descriptores de dicha especie química, y $\{c_j\}$ a un conjunto de coeficientes reales no nulos (excepto quizás c_0), el objetivo supremo de un modelo QSPR es encontrar los valores de los elementos de $\{c_j\}$ y la entidad de los descriptores que hagan que

$$y_i = c_0 + \sum_{j=1}^d c_j d_{ij}$$

sea la *mejor* aproximación global para el conjunto de todas las especies químicas cuyos valores de y pudiera interesar estimar. Evidentemente esto es sólo una expresión de deseo ya que no se dispone de más información que aquella correspondiente a un conjunto limitado de especies químicas.

Comentario 5.3.1. Lo anterior no se condice con la doctrina de considerar importante la obtención de un modelo de aplicabilidad general. Si bien la generalidad de aplicación es deseable, sobre todo desde el punto de vista teórico, en modelos empíricos suele estar asociada con la falta de precisión. Es válida la construcción de modelos, destinada a un subconjunto de casos más o menos bien caracterizado.

El valor de d es establecido por quienes generan el modelo, no debe ser demasiado grande para evitar sobreajuste, lo cual disminuye el poder de predicción del modelo.

Si los descriptores a utilizar en el modelo fueran conocidos, los valores de los elementos de $\{c_j\}$ serían fácilmente obtenibles estableciendo un criterio estándar, por ejemplo, la minimización de los residuos al cuadrado obtenidos con las predicciones, es decir, reducir el problema a un ajuste multivariable por regresión lineal de cuadrados mínimos[42].

El proceso de elección de descriptores constituye el mayor problema de cálculo de la construcción del modelo, por lo que se dedicará la sección 5.3.1 completa a ello.

La aplicación de un modelo está sujeta a su fiabilidad. Por ello, se considera esencial el contar con alguna medida de esta. La calidad de un modelo puede cuantificarse otorgando un valor numérico (a veces llamado puntaje) establecido con alguna función de los residuos obtenidos al emplearlo, por ejemplo la suma de sus cuadrado. Dados dos modelos con l descriptores cada uno, no necesariamente el modelo con mejor calidad en el conjunto de ajuste será el que presente mejor poder predictivo para casos no incluidos en este. Esto es así porque los modelos no fueron ajustados al total de los casos posibles, sino al conjunto de ajuste. Por lo tanto, es inevitable la existencia de algún sesgo. Entonces, el puntaje del modelo en el conjunto de ajuste no es una buena medida de su desempeño en otros conjuntos. Para cuantificar la fiabilidad del modelo puede utilizarse otro conjunto, llamado **conjunto de prueba**.

Una alternativa utilizada para la generación de modelos es subdividir al conjunto de especies químicas usadas anteriormente para el ajuste en dos subconjuntos. Uno de ellos se utiliza en la misma forma que el conjunto original, en este caso para generar varios modelos, y se conserva a aquellos con mejores puntajes. Estos son aplicados sobre el

segundo conjunto, llamado **conjunto de validación**, y se conserva sólo a aquel modelo con mejor puntaje en el mismo. El puntaje del modelo elegido en el conjunto de validación tampoco es una buena medida de su desempeño en otros conjuntos, se espera que confiar en dicho puntaje sea una actitud optimista dado el proceso de su elección. Para cuantificar la calidad de predicción del modelo puede utilizarse al conjunto de prueba.

Existe la posibilidad de validar los modelos sin recurrir a un dado conjunto de prueba. Esto es conocido como *validación cruzada*, y será descrito en la Sección 5.4.

Partición de la base de datos

Resulta importante que los conjuntos no estén sesgados respecto de propiedades de interés. La división del conjunto puede hacerse en forma aleatoria. Normalmente resulta preferible utilizar un algoritmo que asegure la diversidad de compuestos, en especial cuando la cantidad disponible es pequeña.

El método *k-means* permite agrupar n observaciones en k grupos de acuerdo a la semejanza de las mismas. En este contexto la semejanza puede medirse en términos de la distancia euclidiana de los vectores de los descriptores.

Más exactamente, con el método *k-means* se pretende lo siguiente: partir un conjunto de observaciones $\{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n\}$ en k ($k \leq n$) conjuntos \mathcal{S}_i , elementos de $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_k\}$, de acuerdo con

$$\arg \min_{\mathcal{S}} \sum_i^k \sum_{\mathbf{d} \in \mathcal{S}_i} \|\mathbf{d} - \bar{\mathbf{d}}_i\|^2$$

donde $\bar{\mathbf{d}}_i$ es la media de los $\mathbf{d} \in \mathcal{S}_i$. No siempre es posible obtener el mínimo global³.

Un algoritmo de minimización es el siguiente. Dividir las observaciones en los k subconjuntos, por ejemplo, aleatoriamente⁴. Luego, aplicar sucesivamente los siguientes pasos: 1. Calcular los k centroides $\bar{\mathbf{d}}_i$. 2. Reasignar cada observación \mathbf{d}_i al conjunto \mathcal{S}_j cuyo centroide sea más cercano. En la práctica el proceso termina al alcanzar convergencia o un número fijado de iteraciones.

En el contexto de QSAR este método puede utilizarse de la siguiente forma. Partir mediante *k-means* al conjunto total de N moléculas en tantos conjuntos como especies tenga el conjunto de ajuste⁵. Asignar al conjunto de ajuste la molécula más cercana

³Salvo que todas las particiones posibles sean evaluadas.

⁴Para este paso existen diferentes alternativas que no son de interés aquí.

⁵La distancia entre dos moléculas corresponderá a la distancia entre sendos vectores de descriptores.

al centroide de cada subconjunto. Si se desea un conjunto de validación, realizar un procedimiento análogo sobre las moléculas restantes.

5.3.1. Algoritmos de selección de descriptores

En esta sección se describe a algunos de los métodos más usados en la selección de descriptores para la creación de un modelo para aplicación en QSPR/QSAR.

Dado un conjunto compuesto por D descriptores, la cantidad de subconjuntos diferentes no nulos que pueden ser formados a partir del conjunto de origen es

$$\sum_{d=1}^D \binom{D}{d} = \sum_{d=1}^D \frac{D!}{d!(D-d)!}$$

Dado un criterio unívoco de *mejor modelo*, y un conjunto de ajuste único, se obtiene una igualdad entre la expresión anterior y el máximo número de modelos lineales distintos que pueden ser formados con los D descriptores.

En la práctica normalmente se fija el número d de descriptores a incluir en el modelo antes que se ajusten los parámetros correspondientes ($\{c_i\}$). La prueba de todos los modelos posibles de formar, incluso para un dado d comúnmente suele resultar computacionalmente imposible de ser realizada.

Como ejemplo de lo anterior, supóngase que se tienen 3000 descriptores y se desea escoger un conjunto de 8 de ellos, lo cual es un caso razonable en la práctica. Habrán $\approx 1,6 \times 10^{23}$ alternativas posibles. El número de operaciones de coma flotante por segundo realizables por un ordenador personal actual es del orden de 10^2 GFLOPS (del inglés *floating point operations per second*), por esto es en la práctica imposible probar cada posible conjunto de descriptores.

Han sido desarrollados muchos métodos que permiten generar conjuntos de subconjuntos de descriptores, que estadísticamente son más propensos a generar modelos con mejores puntajes, que aquellos formados con los descriptores de los respectivos subconjuntos, de un conjunto de subconjuntos de descriptores creados aleatoriamente. A continuación se describen algunos de estos métodos.

Método de selección “paso a paso”

En este método la inclusión de los descriptores, al conjunto de los descriptores elegidos, se realiza de a uno por vez. En cada etapa de inclusión se agrega al descriptor, cuyo agregado

genere una la mayor mejora estadística al modelo. El proceso termina cuando la inclusión de cualquier nuevo descriptor no genere una mejora estadística significativa.

Una versión alternativa es comenzar con la inclusión de todos los descriptores al modelo, y realizar la eliminación, uno por paso, hasta obtener un deterioro estadístico significativo.

Su popularidad se debe a que es un procedimiento rápido, sencillo, con un costo computacional bajo, y que se halla ampliamente disponible en paquetes computacionales.

Método de reemplazo

El Método de Reemplazo (RM) [43] es un método de selección de variables de bajo costo computacional, que normalmente produce resultados suficientemente cercanos a los de una búsqueda exhaustiva.

El funcionamiento del método es el siguiente. Se escoge, en forma aleatoria, una lista inicial de d descriptores, entre los D disponibles, y se ajusta por cuadrados mínimos los coeficientes de un modelo lineal. En primer instancia se reemplaza al primer descriptor de la lista por los $D - d$ restantes construyéndose $D - d$ modelos. En cada uno de estos modelos, el coeficiente c_i de cada descriptor i entre los $d - 1$ conservados en la lista tomará algún valor. Para cada i , se puede calcular la desviación estándar de conjunto de valores que tomó c_i para en los $D - d$ modelos. Por abuso de lenguaje, a esto se le llamará desviación estándar del descriptor. Dividiendo este valor por la correspondiente media, se obtiene una cantidad que será llamada *desviación estándar relativa* del descriptor (der_i). Se conserva al mejor modelo⁶ entre los $D - d + 1$ obtenidos, y se procede a realizar lo mismo reemplazando a otro descriptor entre los $d - 1$ incluidos. El escogido es aquel con mayor der_i . Si alguno de los nuevos $D - d$ modelos es mejor que el de partida, entonces se realiza el reemplazo del descriptor, y recalculan los nuevos der . Se repite este procedimiento hasta que, sin que se produzca reemplazo alguno, se prueba dos veces el reemplazo del mismo descriptor.

Luego, se realiza todo el procedimiento empezando por cada uno de los $d - 1$ descriptores iniciales restantes. El modelo escogido el de menor desviación estándar entre todos los evaluados.

El RM tiene rápida convergencia, aunque en algunos casos es a mínimos locales no deseables. Este método surgió a partir de la observación empírica, de que las der de los descriptores de los buenos modelos son normalmente chicas. En esta tesis se realizó una breve justificación de el por qué de su funcionamiento.

⁶El criterio suele ser menor desviación estándar de los residuos (S).

Método de reemplazo modificado

El Método de Reemplazo Modificado (MRM)[44, 45], basado en el RM, es más resistente que este último a estancamientos en mínimos locales de bajo poder predictivo.

El algoritmo de este método es similar a aquel del RM, con las siguientes diferencias. En cada paso se sustituye al descriptor, incluso si la sustitución no redujese S . El algoritmo finaliza cuando se cumpla alguna de las siguientes condiciones. Se repite cuatro veces el mismo punto, o se producen 350 pasos. El modelo aceptado es aquel explorado con menor S .

Método de reemplazo ampliado

El método de reemplazo ampliado[46] (ERM, del inglés *Enhanced Replacement Method*) consiste en la aplicación sucesiva y ordenada de los anteriores: RM-MRM-RM. Produce resultados superiores a ambos métodos individuales. A efectos prácticos, su costo computación es ligeramente mayor que el MRM.

5.4. Prueba de modelos

Ya se mencionó que el poder predictivo de un modelo puede medirse utilizando un conjunto de prueba, con moléculas no utilizadas para la creación del modelo.

En ocasiones resulta conveniente cuantificar el poder predictivo de un modelo utilizando sólo las especies del conjunto de ajuste. Por ejemplo, cuando no se dispone de suficientes moléculas para aislar un conjunto de pruebas sin comprometer en gran medida la calidad del ajuste. Incluso, aunque se dispusiera de estas y se validara al modelo con un conjunto de pruebas, puede resultar de interés realizar también validaciones cruzadas.

La validación cruzada consiste hacer múltiples particiones alternativas (ajuste/prueba) del conjunto total de moléculas. En cada una de ellas crear un modelo y probarlo, usando sus respectivos conjuntos de ajuste y prueba. Finalmente se promedian los resultados de las pruebas del modelo, valor que se considera como una medida de la predictibilidad del mismo.

Existen distintos tipos de validación cruzada. Las llamadas *exhaustivas* consisten en utilizar todas particiones posibles donde el conjunto de pruebas contenga (p) de las N moléculas⁷ totales. El número de modelos creados será

⁷Si bien se está hablando en términos de moléculas, estos procedimientos son completamente generales.

$$\binom{N}{p}$$

El método es llamado *Leave-One-Out* (LOO) si $p = 1$. Si $p > 1$ el método es llamado *Leave-More-Out* (L- n -O), donde n representa el porcentaje de moléculas pertenecientes al conjunto de pruebas[47].

Existen también formas de validación cruzada llamadas *no exhaustivas*, en las cuales no se considera a todas las particiones con conjunto de pruebas de p elementos.

Una forma usual de utilizar LOO para la validación es, obtener el coeficiente de determinación (r^2) entre los valores predichos y aceptados de la propiedad. Suele considerarse que el modelo tiene un poder predictivo aceptable cuando $r^2 > 0,5$.

En la referencia [48] los autores proponen que lo anterior no es idóneo y sugieren utilizar, también, las siguientes condiciones como criterio de aceptación del modelo

$$\begin{aligned}\frac{r^2 - r_0^2}{r^2} &< 0,1 \\ \frac{r^2 - r_0'^2}{r^2} &< 0,1 \\ 0,85 &< k < 1,15 \\ 0,85 &< k' < 1,15\end{aligned}$$

donde r_0^2 y $r_0'^2$ representan los coeficientes de determinación para la propiedad observada vs. predicha, y predicha vs. observada, respectivamente, de ajustes lineales con ordenada al origen nula. k y k' representan las respectivas pendientes⁸.

Una prueba que suele hacerse a los modelos es la llamada *y-randomization*[49]. Consiste en permutar aleatoriamente el vector que contiene los valores de la propiedad a predecir. Es decir, los N valores aceptados de la propiedad son asignados aleatoriamente a las N moléculas. Luego, se crea un gran número de modelos de la misma cantidad de descriptores que el modelo original. Se compara luego el poder predictivo de estos últimos modelos con el del modelo original. El modelo original no es confiable si su poder predictivo no es significativamente mejor que los del resto.

Comentario 5.4.1. Este procedimiento tiene mayor utilidad cuando se lo aplica sobre el mismo conjunto de ajuste, comparando las S u otro parámetro estadístico semejante. Nótese que si el modelo es obtenido por ajustes de mínimos cuadrados, entonces, el modelo nunca podrá tener un S superior al que brinda la función

⁸Nótese que r^2 corresponde al ajuste lineal *con* ordenada al origen.

constante igual a la media de las propiedades en el conjunto de ajuste. Más detalladamente, supóngase que se tienen una cantidad muy grande (a efectos prácticos) de moléculas (con estructuras aleatorias) y un número finito de descriptores, con los cuales se pretende crear un modelo de unos pocos descriptores. Se esperaría que de no haber relación alguna entre los descriptores y la propiedad, el coeficiente de cada descriptor sea nulo. Por lo tanto, una regresión lineal multivariable con ordenada al origen, sólo daría un valor no nulo para el término independiente. Puede demostrarse fácilmente que ese valor es la media.

En un caso real los coeficientes de la regresión pueden ser no nulos incluso si no hubiera relación entre los descriptores y la propiedad. Esto podría resultar evidente cuando el número de moléculas es pequeño. El efecto de esto es la obtención de S considerablemente mejores que la media.

La y -randomization permite comparar al modelo con otros modelos que hayan podido ser beneficiados, en la forma recién descrita, por el azar.

5.4.1. Dominio de aplicabilidad

Al proponer modelos lineales, se supone que la respuesta entre los descriptores y la propiedad es aceptablemente representada por una relación lineal. Si sólo se conoce la existencia de una correlación, entonces no hay razón para confiar que esa linealidad se extiende más allá del rango de valores de los descriptores usados para el ajuste.

Se espera entonces que la distancia entre las variables independientes asociadas a una molécula⁹, y la del resto de las moléculas del conjunto de ajuste, no sea muy grande.

Una forma de cuantificar esto es con los *leverages*. El *leverage* h , para la molécula i -ésima (h_i) se define con

$$h_i = \mathbf{x}_i \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{x}_i^T$$

donde \mathbf{x}_i es la matriz de $1 \times d$ que representa a los descriptores de la molécula i , y \mathbf{X} es la matriz de $M \times d$ compuesta por los vectores $\{\mathbf{x}_k\}$ del conjunto de ajuste. Se consideró que existen M moléculas en el conjunto de ajuste y que el modelo contiene d descriptores[50].

Comentario 5.4.2. Conviene realizar algunas aclaraciones. Estos conceptos son más generales que lo recién expuesto. Se llama *matriz sombrero*, o *matriz de influencia*, a una matriz \mathbf{H} que mapea desde el vector de la variable dependiente \mathbf{y} (ejemplo, valores aceptados de la propiedad a predecir), al vector de los valores predichos $\hat{\mathbf{y}}$,¹⁰

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$$

Los elementos diagonales de \mathbf{H} son los *leverages*. Su forma depende del método usado para realizar las predicciones. Si se utiliza regresión lineal de cuadrados mínimos, sin ordenada al origen, con igual peso para todas las variables, y sin correlación entre los errores, se obtiene

⁹En este contexto, descriptores.

¹⁰De allí el nombre “sombrero”.

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

donde \mathbf{X} es la matriz de a variable explicativa (definida anteriormente en el texto). De aquí que los *leverages* de los descriptores se definieran como se lo hizo.

Esta expresión puede resultar poco útil a efectos interpretativos. En una regresión lineal simple, es bastante directo obtener mediante la expresión dada para \mathbf{H} , que los elementos diagonales tienen la forma

$$h_{ii} = \frac{\mathbf{X}_i^2}{\sum_j^n \mathbf{X}_j^2}$$

Si se aplicara una traslación que ubique al origen de coordenadas en el valor promedio de \mathbf{X} ($\bar{\mathbf{X}}$), es decir $\tilde{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{X}}$, se obtiene

$$h_{ii} = \frac{(\mathbf{X}_i - \bar{\mathbf{X}}_i)^2}{\sum_j^n (\mathbf{X}_j - \bar{\mathbf{X}}_j)^2}$$

Este resultado es interesante aquí, en tanto que evidencia para este caso, que los leverages son una medida de cuan lejano es un conjunto de descriptores, del promedio de los utilizados para la realización del ajuste.

En caso que se incluya la ordenada al origen, a la matriz \mathbf{X} se le agrega una columna de unos, e incluso sin el centrado al valor promedio se obtiene

$$h_{ii} = \frac{1}{n} + \frac{(\mathbf{X}_i - \bar{\mathbf{X}}_i)^2}{\sum_j^n (\mathbf{X}_j - \bar{\mathbf{X}}_j)^2}$$

Esto tiene relación directa con la distancia de Mahalanobis, definida con

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$$

donde \mathbf{S} es la matriz de covarianza,

$$d(\mathbf{x}, \bar{\mathbf{x}})^2 = (n-1) \left(h_{ii} + \frac{1}{n} \right)$$

Donde el término $1/n$ sólo en caso de ajustes con ordenada al origen.

Los leverages están acotados, $0 \leq h_{ii} \leq 1$, y se cumple

$$\sum_i h_{ii} = d$$

donde d representa al número de coeficientes del modelo (aquí descriptores).

No existe una definición rigurosa de *grande* para un leverage, pero normalmente se utiliza $3d/n$ donde n es el número de observaciones (aquí: número de moléculas del conjunto de ajuste)[51].

Si el leverage de una molécula sobre la cual se realiza una predicción es grande, esta no es confiable. Esto no quiere decir que sea mala. Si el modelo fuese de un único descriptor,

con la forma $y_i = a x_i$, esto implicaría un y_i alejado del de las moléculas del conjunto de ajuste¹¹. A lo que se llama valor atípico (o *outlier* en inglés). Pero esto no es necesariamente así si el modelo incluye a más de un descriptor.

Se espera que el conjunto de ajuste no esté sesgado. Para tal fin es conveniente que el conjunto tenga especies con propiedades diversas, pero también que no haya especies que produzcan un sesgo del modelo debido a que ejerzan una gran *influencia* en el ajuste.

Los outliers y moléculas con grandes leverages son potencialmente influyentes. Un outlier que mantenga la tendencia¹² no será influyente¹³. Por ejemplo, una especie cuyo x_i lejano al resto permita una buena predicción de su propiedad mediante $y_i = a x_i$. En el caso multidimensional esto también es válido. Un outlier que no siga la tendencia puede o no ser influyente. En el caso unidimensional del ejemplo, la mayor influencia se da para outliers con grandes leverages.

El dominio de aplicabilidad puede considerarse una región en un espacio d -dimensional. En cualquier caso, para visualizarlo puede usarse el gráfico de Williams. En él se grafican los *residuos estandarizados* (σ_i), es decir, los residuos (res_i) divididos entre la desviación estándar de los residuos en el conjunto de ajuste (S_{tr})

$$\sigma_i = \frac{\text{res}_i}{S_{tr}}$$

¹¹Se está considerando que en el conjunto de ajuste no hay moléculas con leverages grandes.

¹²Aquí tendencia hace referencia a un comportamiento descrito por un modelo que describa bien al resto de los casos.

¹³Porque se ve bien representado por el modelo.

Capítulo 6

Algoritmos desarrollados para la construcción de modelos QSPR

6.1. Variación sobre el método reemplazo modificado

Un objetivo principal de esta tesis es la creación de descriptores. Para ello se requiere utilizar un método de selección de descriptores. Uno de los métodos más eficientes para tal fin es el método de reemplazo (ver sección 5.3.1). Este método fue creado en base a la observación de que a los modelos obtenidos haciendo una búsqueda exhaustiva les corresponde desviaciones relativas chicas de los coeficientes de ajustes multilineales.

Dado que este método puede converger a soluciones que disten significativamente de la óptima, se han creado diferentes modificaciones del mismo, entre las que se describió al Método de Reemplazo Modificado (ver sección 5.3.1) propuesta por el Dr. Andrew Mercader.

Aunque el MRM satisface todos los requerimientos para ser utilizado en este trabajo de tesis, resultó un deseo personal realizar un aporte en esa dirección. Con el fin de comprender el por qué del buen funcionamiento del MR, en primera instancia se analizó e interpretó el efecto asociado a las desviaciones estándar relativas que sirvió de base para este y sus derivados.

6.1.1. Mínimos cuadrados, proyecciones y el método de reemplazo

Represente \mathcal{V} un espacio vectorial de dimensión finita equipado con un producto interno $\langle \cdot, \cdot \rangle$.

Sea $\mathcal{S} \subset \mathcal{V}$. Se llama *complemento ortogonal* de \mathcal{S} al conjunto

$$\mathcal{S}^\perp := \{v \in \mathcal{V} : v \perp s \forall s \in \mathcal{S}\}$$

Si $\mathcal{V} = \mathcal{S} \oplus \mathcal{S}^\perp$, entonces $v = s + s^\perp \forall v \in \mathcal{V}, s \in \mathcal{S}, s^\perp \in \mathcal{S}^\perp$, se llama *proyección ortogonal* sobre el subespacio \mathcal{S} al mapa

$$\mathcal{P}_{\mathcal{S}}(v) = s$$

Análogamente se puede definir a la proyección ortogonal sobre \mathcal{S}^\perp con

$$\mathcal{P}_{\mathcal{S}^\perp}(v) = s^\perp$$

Si se define la adición de operadores en la forma usual, resultará que

$$\mathcal{I} = \mathcal{P}_{\mathcal{S}} + \mathcal{P}_{\mathcal{S}^\perp}$$

donde \mathcal{I} representa al operador identidad.

Represente $y \in \mathcal{V}$ un vector cuyas componentes representen los valores verdaderos de la propiedad a predecir, y $\{x_1, x_2, \dots, x_n\}$ a un conjunto de vectores cuyas componentes correspondan a los respectivos valores para las distintas especies químicas del conjunto de ajuste.

Puede demostrarse que el ajuste de regresión lineal corresponde a la proyección de y sobre el *espacio generado* por $\{x_1, x_2, \dots, x_n\}$, $\mathcal{X} := \text{span}(\{x_1, x_2, \dots, x_n\})$. Si se representa con $\mathcal{P}_{\mathcal{X}}$ al correspondiente operador de proyección, y con \tilde{y} al vector cuyos valores son los correspondientes al ajuste, se cumple

$$\tilde{y} = \mathcal{P}_{\mathcal{X}} y$$

Esto es relevante porque hace evidente que el vector de los residuos $(y - \tilde{y})$ es ortogonal a \mathcal{X} , ya que

$$y - \tilde{y} = (\mathcal{I} - \mathcal{P}_{\mathcal{X}})(y) = \mathcal{P}_{\mathcal{X}^\perp}(y)$$

y $\mathcal{P}_{\mathcal{X}^\perp} := \mathcal{I} - \mathcal{P}_{\mathcal{X}}$ representa al operador que proyecta en un subespacio ortogonal a \mathcal{X} .

La ortogonalización puede utilizarse para encontrar los coeficientes la regresión multilíneal. Para hacer una regresión de y con $\{x_1, x_2, \dots, x_n\}$, puede realizarse un procedimiento de n pasos, uno por cada x_i , que consistan en realizar las 3 etapas siguientes:

1. $z_1 := x_1$

2. $\forall j \in \mathbb{N} \leq n$, calcular $z_{j>1} := x_j - \sum_{i=1}^{j-1} \frac{\langle z_i, x_j \rangle}{\|z_i\|^2} z_i$

3. Hacer la regresión lineal de y con z_n , para obtener el coeficiente c_n :

$$c_n = \frac{\langle z_n, y \rangle}{\|z_n\|^2}$$

Puede demostrarse que el coeficiente c_n también corresponde al de la variable x_n la regresión lineal de y frente a $\{x_1, x_2, \dots, x_n\}$. Nótese que el orden de las variables no tiene importancia.

Como cada z_i fue definido como el residuo de la regresión de x_i frente a z_1, z_2, \dots, z_{i-1} , los z_k son ortogonales entre sí.

Para comprender la base del funcionamiento del MR y sus derivados es clave notar que el coeficiente del vector x_n en la regresión multilineal depende en gran medida de la similitud entre x_n y $\{x_{i \neq n}\}$. Esto se ejemplificará brevemente a continuación.

A modo ilustrativo considérese el caso extremo en el cual $\langle x_n, x_i \rangle = 0 \forall i \neq n$, que implicaría $z_n = x_n$ y por lo tanto el coeficiente de la regresión multilineal de x_n sería el mismo que aquel de la regresión lineal de esta variable aislada. El caso extremo opuesto ($\langle x_n, x_i \rangle = \|x_n\| \|x_i\| \forall i \neq n$) no es de interés dado que genera una singularidad.

Observe que z_n es el residuo de realizar una regresión lineal del vector x_n frente a $\{x_{i \neq n}\}$. En caso que haya al menos algún x_m con dependencia *quasi*-lineal con x_n , se obtendrá $\|z_n\| \approx 0$, haciendo z_n (y z_m) muy sensible a la variación de los $\{x_{i \neq m, n}\}$.

En lo anterior radica la esencia del funcionamiento del MR y sus derivados. Durante cada paso del proceso se realiza la sustitución de algún $x_{1 \leq k \leq n}$ por el resto de los $x_{j > n}$, y los coeficientes de las regresiones lineales, de aquellos $\{x_{i \neq k}\}$ con mayor dependencia lineal entre sí, tenderán a variar en mayor medida y aquel con mayor desviación media relativa será electo como candidato a ser reemplazado.

Una de las bondades del MR es entonces la tendencia a la generación de modelos con poca dependencia lineal entre sus variables, sin requerir una previa reducción, debido a dependencias *quasi*-lineales, de variables potenciales. Esto es particularmente bueno dado que se reduce la redundancia.

Desafortunadamente existen al menos dos problemas inherentes al mismo mecanismo de elección. Uno de ellos es que la tendencia de una variable a ser reemplazada aumenta con la dependencia lineal entre ella y las variables no seleccionadas para el modelo. Esto es desafortunado dado que de dichas dependencias nada puede deducirse acerca del poder predictivo del modelo.

El segundo problema ocurre cuando se escogen variables con poco poder predictivo que tengan una relativamente alta independencia lineal con el resto. En cuyo caso se obtendrá un modelo final con poco poder predictivo. Este último problema es de mayor interés práctico y aquel que ha sido atacado desarrollando el MRM.

6.1.2. Variantes consideradas

Luego de haber analizado las bases del funcionamiento del MR, se exploró algunas variantes del mismo. Para ello se utilizó una base de datos de 166 compuestos orgánicos (y sus 1268 descriptores calculados con Dragon)[52] de los que se sabe sus fluorofilicidad y se implementaron las mismas en Python3. Se realizaron 500 corridas con cada variante, descriptas a continuación. Los resultados son reportados en la tabla 6.1.

La variante más directa del MR considerada (a la que se referirá con v_1) consiste en aceptar el reemplazo del descriptor siempre que la desviación estándar (S), de los errores que genere el cambio, no supere $x\hat{S}$ ($x>1$) veces la menor S alcanzada durante el proceso. El proceso finaliza cuando no sea posible el reemplazo. El conjunto de descriptores aceptado será aquel que haya dado la menor S . Para el ensayo se utilizó $x = 1,05$.

Con el fin de seguir con el mismo esquema de los métodos anteriores, en cada paso se asignará a cada descriptor presente en el modelo preliminar (excepto aquel que está siendo variado) un puntaje que será proporcional a la desviación estándar relativa del coeficiente (al igual que en el MR y sus derivados), pero también lo será a una función monótona del error medio (\bar{e}) de las estimaciones obtenidas a partir de la regresión lineal del vector que representa al descriptor frente a aquel que se quiere predecir. Con esto último se consigue aumentar la probabilidad de que un descriptor sea elegido como candidato a ser reemplazado al disminuir su aporte a la propiedad de interés. La función elegida fue $f(\bar{e}) = \bar{e}$. En otras palabras, en vez de elegir al nuevo descriptor a reemplazar por su mayor desviación estándar relativa, se lo elige por el producto entre esta y \bar{e} . Se llamará a esta variación v_2 .

Se consideró también a los dos casos extremos. Uno consiste en probar el reemplazo de todos los descriptores temporalmente incluidos en el modelo. A esto se le llamará FRM. El otro caso extremo es aquel en el que se considera un único descriptor elegido al azar. A este caso se lo denotará por RRM.

Los distintos orígenes de los caminos recorridos por el MR y sus derivados son en principio equivalentes. Las comparaciones entre las variantes mencionadas pueden realizarse entonces utilizando un único camino. De esta forma las diferencias debieran verse acentuadas. Las comparaciones se hicieron utilizando un único camino por método. Los

Método	Media	Desv. Estándar	Min.	Máx.
RM	0.5308	0.0414	0.4452	0.7416
ν_1	0.5293	0.0394	0.4593	0.7240
ν_2	0.5282	0.0329	0.4531	0.6811
RRM	0.5785	0.0686	0.4574	0.9559
FRM	0.5035	0.0323	0.4176	0.6017

Cuadro 6.1 – Resultados de error medio, desviación estándar de errores, errores mínimo máximo para los distintos métodos de esta sección. La nomenclatura está descrita en el texto principal.

resultados de la tabla 6.1 inducen a pensar que el MR tiene un desempeño significativamente mejor que el RRM, e inferior al del FRM. Dado que el costo computacional del MR es similar al de RRM, pero que el FRM es tantas veces más costoso que el MR como descriptores tenga el modelo, las ventajas del MR resultan claras. Con los datos obtenidos no puede concluirse la superioridad del método ν_1 por sobre el RM. El método ν_2 es más robusto que el del MR, aunque su costo computacional es ligeramente mayor. Una comparación más minuciosa sería necesaria para comparar el costo/beneficio entre estos métodos.

Parte IV

METODOLOGÍA

Capítulo 7

Bases de datos

Se desea estudiar potenciales correlaciones entre los descriptores desarrollados en esta tesis y propiedades experimentales. Para eso se utilizaron tres bases de datos. Una de ellas contiene 160 moléculas. Sus coeficientes de partición entre octanol y agua ($\log P$) a 298.15 K son conocidos. La misma es un subconjunto de la base de datos descrita en Schnackenberg[53]. La remoción de dos moléculas fue realizada con el fin de posibilitar la comparación de los descriptores calculados el programa eDragon[54], ya que este falló con dichas moléculas.

La tabla 7.1 contiene la lista de nombres y sus correspondientes estructuras en formato SMILES (*Simplified Molecular Input Line Entry Specification*). Se utilizó la lengua inglesa para nombrar los compuestos porque esto simplifica el proceso de búsqueda de las mismas en bases de datos sustancias. El significado preciso del superíndice TS está descrito en la sección de resultados.

Nombre	SMILE	Nombre	SMILE
Naphthalen-2-amine	<chem>Nc1cc2c(cc1)cccc2</chem>	Phenylmethanamine	<chem>NCc1ccccc1</chem>
Acetic acid ^{TS}	<chem>CC(=O)O</chem>	Benzamide	<chem>O=C(N)c1ccccc1</chem>
1H-Imidazole ^{TS}	<chem>n1cc[nH]c1</chem>	N-Methylbenzamide	<chem>O=C(NC)c1ccccc1</chem>
Propanoic acid ^{TS}	<chem>OC(=O)CC</chem>	N-Methylaniline ^{TS}	<chem>N(c1ccccc1)C</chem>
Aniline ^{TS}	<chem>c1(ccccc1)N</chem>	Phenylcyanamide ^{TS}	<chem>N(c1ccccc1)C#N</chem>
Phenylmethanol	<chem>OCc1ccccc1</chem>	N-Phenylacetamide	<chem>O=C(Nc1ccccc1)C</chem>
2-Hydroxybenzamide	<chem>Oc1c(cccc1)C(=O)N</chem>	N-Phenylhydroxylamine ^{TS}	<chem>ONc1ccccc1</chem>
4-Hydroxybenzoic acid	<chem>Oc1ccc(cc1)C(=O)O</chem>	N-Phenylaniline	<chem>N(c1ccccc1)c1ccccc1</chem>
2,4-Dihydroxybenzoic acid	<chem>c1(cc(ccc1C(=O)O)O)O</chem>	Phenol	<chem>c1(ccccc1)O</chem>
Benzoic acid	<chem>OC(=O)c1ccccc1</chem>	ethanol	<chem>CCO</chem>
2-(3-methylphenyl)acetic acid	<chem>OC(=O)Cc1cc(ccc1)C</chem>	Methanamine	<chem>CN</chem>
3-Methoxybenzoic acid	<chem>COc1cccc(c1)C(=O)O</chem>	Methanol ^{TS}	<chem>CO</chem>
N,N-dihydroxy-4-methylaniline	<chem>O=N(=O)c1ccc(cc1)C</chem>	Octan-1-ol	<chem>C(CCCCCO)CCC</chem>
4-Phenylbutanoic acid	<chem>C(Cc1ccccc1)CC(=O)O</chem>	Ethanamine	<chem>NCC</chem>
2-Methylbenzoic acid ^{TS}	<chem>OC(=O)c1c(C)cccc1</chem>	N-Methylmethanamine ^{TS}	<chem>N(C)C</chem>
Naphthalen-1-ol	<chem>Oc1c2c(ccc1)cccc2</chem>	Cyclohexanol	<chem>OC1CCCCC1</chem>
Di(phenyl)methanone	<chem>O=C(c1ccccc1)c1ccccc1</chem>	N-ethyl-1-phenylcyclohexan-1-amine	<chem>N(C1(CCCCC1)c1ccccc1)CC</chem>
Naphtalene	<chem>c1cccc2ccccc12</chem>	2H-triazole	<chem>c1cn[nH]n1</chem>
4-Butylbenzoic acid	<chem>OC(=O)c1ccc(CCCC)cc1</chem>	1H-1,2,4-Triazole ^{TS}	<chem>n1cn[nH]c1</chem>
1,2,3,4,5-Pentamethylbenzene	<chem>Cc1c(C)c(C)c(cc1C)C</chem>	4-Nitroaniline	<chem>O=N(=O)c1ccc(N)cc1</chem>
2-phenylethylbenzene	<chem>C(Cc1ccccc1)c1ccccc1</chem>	3-Nitroaniline	<chem>O=N(=O)c1cc(N)ccc1</chem>
2-Methylphenanthrene	<chem>c1(C)cc2c(cc1)c1c(cc2)cccc1</chem>	2-Nitroaniline ^{TS}	<chem>O=N(=O)c1c(N)cccc1</chem>
Pyrene	<chem>c12c3c4ccc2cccc1ccc3ccc4</chem>	Benzene-1,4-diol	<chem>Oc1ccc(cc1)O</chem>
2-Ethylanthracene	<chem>C(C)c1cc2c(cc1)cc1c(c2)cccc1</chem>	Benzene-1,3-diol ^{TS}	<chem>Oc1cc(O)ccc1</chem>
Benzene ^{TS}	<chem>c1ccccc1</chem>	benzene-1,2-diol	<chem>Oc1c(O)cccc1</chem>
Methylbenzene ^{TS}	<chem>c1(ccccc1)C</chem>	4-Nitrophenol	<chem>c1c(ccc(c1)N(=O)=O)O</chem>
Ethylbenzene	<chem>CCc1ccccc1</chem>	3-Nitrophenol	<chem>Oc1cc(N(=O)=O)ccc1</chem>
Ethenylbenzene	<chem>c1(ccccc1)C=C</chem>	2-(Dihydroxyamino)phenol ^{TS}	<chem>c1(ccccc1O)N(=O)=O</chem>
Ethynylbenzene ^{TS}	<chem>C#Cc1ccccc1</chem>	4-Nitrobenzoic acid	<chem>O=N(=O)c1ccc(C(=O)O)cc1</chem>
Propylbenzene	<chem>C(Cc1ccccc1)C</chem>	3-Nitrobenzoic acid ^{TS}	<chem>OC(=O)c1cc(N(=O)=O)ccc1</chem>
Prop-2-enylbenzene	<chem>C(=C)Cc1ccccc1</chem>	2-Nitrobenzoic acid ^{TS}	<chem>OC(=O)c1c(N(=O)=O)cccc1</chem>
Isopropylbenzene	<chem>CC(C)c1ccccc1</chem>	4-Aminobenzoic acid ^{TS}	<chem>Nc1ccc(cc1)C(=O)O</chem>
Butylbenzene ^{TS}	<chem>C(Cc1ccccc1)CC</chem>	4-Aminophenol	<chem>Nc1ccc(O)cc1</chem>
Tert-Butylbenzene	<chem>c1cc(ccc1)C(C)(C)C</chem>	4-Cyanobenzoic acid ^{TS}	<chem>OC(=O)c1ccc(cc1)C#N</chem>
Cyclopropylbenzene	<chem>C1C(C1)c1ccccc1</chem>	4-Hydroxybenzonitrile ^{TS}	<chem>Oc1ccc(cc1)C#N</chem>
Phenylbenzene	<chem>c1cc(ccc1)c1ccccc1</chem>	4-Methoxybenzoic acid	<chem>OC(=O)c1ccc(cc1)OC</chem>
phenylmethylbenzene	<chem>C(c1ccccc1)c1ccccc1</chem>	2-Methylpyridine	<chem>n1c(cccc1)C</chem>
Ethane	<chem>CC</chem>	2-Pyridinecarbonitrile ^{TS}	<chem>c1ccc(nc1)C#N</chem>
Ethene	<chem>C=C</chem>	Ethyl pyridine-2-carboxylate	<chem>O(C(=O)c1cccn1)CC</chem>
Acetylene	<chem>C#C</chem>	Methyl pyridine-2-carboxylate	<chem>O(C(=O)c1cccn1)C</chem>
Propane ^{TS}	<chem>C(C)C</chem>	N,N-Dimethylpyridin-2-amine	<chem>N(c1ncccc1)(C)C</chem>
Prop-1-ene ^{TS}	<chem>CC=C</chem>	2-Ethoxypyridine ^{TS}	<chem>C(Oc1cccn1)C</chem>
Prop-1-yne	<chem>CC#C</chem>	2-Methoxypyridine	<chem>COc1cccn1</chem>
Butane	<chem>C(CC)C</chem>	1-(2-Pyridyl)ethanone	<chem>CC(=O)c1cccn1</chem>
2-Methylpropane	<chem>C(C)(C)C</chem>	Pyridin-2-ylmethyl acetate	<chem>O(Cc1ncccc1)C(=O)C</chem>
Buta-1,3-diene	<chem>C(=C)C=C</chem>	2-Ethylpyridine ^{TS}	<chem>n1c(CC)cccc1</chem>
Pentane	<chem>C(CC)CC</chem>	2-Phenylpyridine	<chem>c1ccc(nc1)c1ccccc1</chem>
2,2-Dimethylpropane	<chem>C(C)(C)(C)C</chem>	Pyrazine ^{TS}	<chem>n1ccncc1</chem>
Cyclopropane ^{TS}	<chem>C1CC1</chem>	2-Methylpyrazine	<chem>n1c(C)cncc1</chem>

Cyclopentane	C1CCCC1	Pyrazine-2-carbonitrile	c1c(nccn1)C#N
Cyclohexane	C1CCCCC1	Ethyl pyrazine-2-carboxylate	O(CC)C(=O)c1ncnc1
1,2-Dimethylbenzene	Cc1c(C)cccc1	Methyl pyrazine-2-carboxylate ^{TS}	c1nc(cnc1)C(=O)OC
1,3-Dimethylbenzene	Cc1cc(C)ccc1	3,5-Dimethylpyrazin-2-amine	n1c(c(nc1C)N)C
1,4-Dimethylbenzene ^{TS}	Cc1ccc(cc1)C	2-Ethoxypyrazine ^{TS}	c1(cncn1)OCC
1,2,3,4,5,6-hexamethylbenzene	c1(C)c(C)c(C)c(c(c1C)C)C	2-Methoxypyrazine	COc1ncnc1
1,2,3,4-Tetrahydronaphthalene	c12c(CCCC2)cccc1	1-Pyrazin-2-ylethanone ^{TS}	c1c(nccn1)C(=O)C
Furan ^{TS}	o1cccc1	2-Ethylpyrazine	n1c(cnc1)CC
1-Phenylpyrrole	clcccc(c1)n1cccc1	2-Phenylpyrazine	n1c(c2cccc2)cncc1
Anthracene	c12c(cc3c(c1)cccc3)cccc2	Pyrimidine	c1cncnc1
Phenanthrene	clcc2c(cc1)ccc1c2cccc1	2-Methylpyrimidine ^{TS}	n1c(ncnc1)C
Benzonitrile	clcccc(c1)C#N	Pyrimidine-2-carbonitrile ^{TS}	n1ccnc1C#N
1-Phenylethanone	O=C(c1cccc1)C	N,N-dimethylpyrimidin-2-amine	N(c1ncnc1)(C)C
Ethyl benzoate	O(CC)C(=O)c1cccc1	2-Ethoxypyrimidine	O(CC)c1ncnc1
Methyl benzoate ^{TS}	COC(=O)c1cccc1	2-Methoxypyrimidine ^{TS}	COc1ncnc1
N,N-dimethylaniline	N(c1cccc1)(C)C	4-Methylpyrimidine ^{TS}	n1c(C)ccnc1
Nitrobenzene ^{TS}	O=N(=O)c1cccc1	Oxirane	O1CC1
Ethoxybenzene	C(Oc1cccc1)C	Pyridine ^{TS}	n1cccc1
Methoxybenzene	COc1cccc1	Quinoline ^{TS}	clccc2ncccc2c1
Acetonitrile	CC#N	Methyl acetate	CC(=O)OC
Propan-2-one	C(=O)(C)C	Ethyl propanoate	O(C(=O)CC)CC
Methoxymethane	O(C)C	Pentan-2-one	CCCC(=O)C
Cyclohexanone	O=C1CCCCC1	Propanenitrile	C(#N)CC9
1,2-oxazole	clcnol	Ethyl-dihydroxyazanium	O=N(=O)CC
Benzaldehyde	O=Cc1cccc1	4-Ethoxypyrimidine	O(CC)c1ncnc1
Propanal	CCC=O	5-Methylpyrimidine	n1cc(C)cnc1
2-Phenylacetonitrile ^{TS}	clcccc(c1)CC#N	Pyrimidine-5-carbonitrile	n1cc(cnc1)C#N
1-phenylpropan-2-one ^{TS}	O=C(Cc1cccc1)C	Ethyl pyrimidine-5-carboxylate	O(CC)C(=O)c1cncnc1
N,N-dihydroxymethanamine	CN(=O)=O	Eethyl pyrimidine-5-carboxylate	O(C(=O)c1cncnc1)C
2-Phenylacetamide ^{TS}	O=C(N)Cclcccc1	5-Ethoxypyrimidine ^{TS}	O(CC)c1cncnc1
2-Phenylacetic acid	C(=O)(Cc1cccc1)O	5-Methoxypyrimidine ^{TS}	O(c1cncnc1)C

Cuadro 7.1 – Nombres y SMILES de las moléculas de la base de datos de 160 compuestos.

La segunda base de datos usada contiene 276 compuestos orgánicos cuyas temperaturas de ebullición son conocidas. La misma corresponde a la base de datos descrita en Schnackenberg[55] luego de que una especie sea removida. La remoción se debió a un error de notación de la fuente original.

Los nombres de las especies y sus correspondientes SMILES se encuentran en la Tabla 7.2.

Nombre	SMILE	Nombre	SMILE
bromotrichloromethane	<chem>BrC(Cl)(Cl)Cl</chem>	3-methyl-2-butanol	<chem>CC([C@H](C)O)C</chem>
chlorotrifluoromethane	<chem>C(Cl)(F)(F)F</chem>	methyl n-butyl ether	<chem>CCCCOC</chem>
carbon tetrachloride	<chem>C(Cl)(Cl)(Cl)Cl</chem>	methyl sec-butyl ether	<chem>O([C@@H](CC)C)C</chem>
carbon tetrafluoride	<chem>C(F)(F)(F)F</chem>	methyl tert-butyl ether	<chem>O(C(C)(C)C)C</chem>
tribromomethane	<chem>BrC(Br)Br</chem>	1-pentanol	<chem>OCCCCC</chem>
chlorodifluoromethane	<chem>ClC(F)F</chem>	2-pentanol	<chem>C[C@H](O)CCC</chem>
dichlorofluoromethane	<chem>ClC(Cl)F</chem>	3-pentanol	<chem>OC(CC)CC</chem>
chloroform	<chem>C(Cl)(Cl)Cl</chem>	1,5-pentanediol	<chem>OCCCCCO</chem>
trifluoromethane	<chem>C(F)(F)F</chem>	methyl n-butyl sulfide	<chem>CCCCSC</chem>
dibromomethane	<chem>BrCBr</chem>	methyl tert-butyl sulfide	<chem>S(C(C)(C)C)C</chem>
dichloromethane	<chem>C(Cl)Cl</chem>	n-pentyl mercaptan	<chem>SCCCCC</chem>
difluoromethane	<chem>C(F)F</chem>	hexachlorobenzene	<chem>Cl(c(c(c(c(Cl)Cl)Cl)Cl)Cl)Cl</chem>
diiodomethane	<chem>ICI</chem>	hexafluorobenzene	<chem>Fc1c(F)c(F)c(F)c(F)c1F</chem>
formaldehyde	<chem>O=C</chem>	1,2,4-trichlorobenzene	<chem>Clc(c(cc(Cl)Cl)Cl)Cl</chem>
formic acid	<chem>C(=O)O</chem>	m-dichlorobenzene	<chem>Clc1cc(Cl)ccc1</chem>
methyl bromide	<chem>BrC</chem>	o-dichlorobenzene	<chem>Clc1c(Cl)cccc1</chem>
methyl chloride	<chem>CCl</chem>	p-dichlorobenzene	<chem>Clc(ccc(Cl)Cl)Cl</chem>
methyl fluoride	<chem>CF</chem>	m-chlorophenol	<chem>Clcc(cc(Cl)Cl)O</chem>
methyl iodide	<chem>IC</chem>	o-chlorophenol	<chem>Clc(c(ccc1Cl)O)</chem>
methanol	<chem>CO</chem>	p-chlorophenol	<chem>Clc(ccc(Cl)Cl)O</chem>
methyl mercaptan	<chem>CS</chem>	phenol	<chem>Cl(ccccc1O)</chem>
1,1-dichlorotetrafluoroethane	<chem>C(C(Cl)(Cl)F)(F)(F)F</chem>	1,2-benzenediol	<chem>Oc1c(O)cccc1</chem>
hexafluoroethane	<chem>FC(F)(F)C(F)(F)F</chem>	1,3-benzenediol	<chem>Oc1cc(O)ccc1</chem>
1,1-dichloro-2,2,2-trifluoroethane	<chem>C(C(F)(F)F)(Cl)Cl</chem>	p-hydroquinone	<chem>Oc1ccc(cc1)O</chem>
trichloroethylene	<chem>C(=C(Cl)Cl)Cl</chem>	cyclohexanone	<chem>O=C1CCCCC1</chem>
trichloroacetaldehyde	<chem>ClC(Cl)(Cl)C=O</chem>	ethyl acetoacetate	<chem>C(C(=O)OCC)C(=O)C</chem>
pentachloroethane	<chem>ClC(Cl)(Cl)C(Cl)Cl</chem>	diethyl oxalate	<chem>O(CC)C(=O)C(=O)OCC</chem>
pentafluoroethane	<chem>FC(F)(F)C(F)F</chem>	cyclohexanol	<chem>OC1CCCCC1</chem>
1,1,2,2-tetrabromoethane	<chem>BrC(Br)C(Br)Br</chem>	3,3-dimethyl-2-butanone	<chem>O=C(C(C)(C)C)C</chem>
1,1,1,2-tetrachloroethane	<chem>ClC(Cl)(CCl)Cl</chem>	ethyl isopropyl ketone	<chem>O=C(C(C)C)CC</chem>
1,1,2,2-tetrachloroethane	<chem>ClC(Cl)C(Cl)Cl</chem>	1-hexanal	<chem>O=CCCCCC</chem>
1,1-difluoroethylene	<chem>FC(=C)F</chem>	2-hexanone	<chem>O=C(CCCC)C</chem>
1,1,2,2-tetrafluoroethane	<chem>C(C(F)F)(F)F</chem>	3-hexanone	<chem>O=C(CCC)CC</chem>
vinyl bromide	<chem>BrC=C</chem>	methyl isobutyl ketone	<chem>C(C(C)C)C(=O)C</chem>
acetyl chloride	<chem>C(=O)(C)Cl</chem>	3-methyl-2-pentanone	<chem>O=C([C@@H](CC)C)C</chem>
1,1,1-trichloroethane	<chem>CC(Cl)(Cl)Cl</chem>	n-butyl acetate	<chem>O(CCCC)C(=O)C</chem>
1,1,2-trichloroethane	<chem>ClC(Cl)CCl</chem>	sec-butyl acetate	<chem>O([C@H](CC)C)C(=O)C</chem>
vinyl fluoride	<chem>C(=C)F</chem>	tert-butyl acetate	<chem>O(C(C)(C)C)C(=O)C</chem>
1,1,1-trifluoroethane	<chem>CC(F)(F)F</chem>	ethyl n-butyrate	<chem>C(C(=O)OCC)CC</chem>
1,1-dibromoethane	<chem>BrC(Br)C</chem>	2-ethyl butyric acid	<chem>OC(=O)C(CC)CC</chem>
1,2-dibromoethane	<chem>BrCCBr</chem>	ethyl isobutyrate	<chem>O(C(=O)C(C)C)CC</chem>
1,1-dichloroethane	<chem>C(C)(Cl)Cl</chem>	n-hexanoic acid	<chem>CCCCC(=O)O</chem>
1,2-dichloroethane	<chem>C(CCl)Cl</chem>	isobutyl acetate	<chem>C(OC(=O)C)C(C)C</chem>
1,1-difluoroethane	<chem>FC(F)C</chem>	n-pentyl formate	<chem>O(CCCCC)C=O</chem>
1,2-difluoroethane	<chem>FCCF</chem>	n-propyl propionate	<chem>CCCOC(=O)CC</chem>
acetic acid	<chem>CC(=O)O</chem>	n-butyl ethyl ether	<chem>O(CCCC)CC</chem>
methyl formate	<chem>COC=O</chem>	tert-butyl ethyl ether	<chem>O(C(C)(C)C)CC</chem>
bromoethane	<chem>CCBr</chem>	diisopropyl ether	<chem>O(C(C)C)C(C)C</chem>
ethyl chloride	<chem>CCCl</chem>	2-ethyl-1-butanol	<chem>OCC(CC)CC</chem>

ethyl fluoride	CCF	1-hexanol	C(CCCO)CC
ethyl iodide	CCI	2-hexanol	O[C@@H](CCCC)C
dimethyl ether	O(C)C	2-methyl-1-pentanol	OC[C@@H](CCC)C
ethanol	CCO	4-methyl-2-pentanol	C(C(C)C)[C@H](C)O
ethylene glycol	C(CO)O	methyl n-pentyl ether	O(CCCCC)C
dimethyl sulfide	S(C)C	1,6-hexanediol	OCCCCCO
ethyl mercaptan	CCS	hexylene glycol	C(C(C)(C)O)[C@H](C)O
dimethyl disulfide	S(SC)C	di-n-propyl sulfideb	S(CCC)CCC
1,2-ethanedithiol	C(CS)S	ethyl tert-butyl sulfide	S(C(C)(C)C)CC
2-chloropropene	ClC(=C)C	n-hexyl mercaptan	C(CCCS)CC
1,2,3-trichloropropane	C(CCl)(CCl)Cl	di-n-propyl disulfide	S(SCCC)CCC
1,2-dichloropropane	C(Cl)[C@H](Cl)C	benzotrichloride	c1cc(ccc1)C(Cl)(Cl)Cl
acetone	C(=O)(C)C	benzyl dichloride	ClC(Cl)c1ccccc1
1-propanal	CCC=O	benzoic acid	OC(=O)c1ccccc1
1,2-propyleneoxide	O1[C@H](C)C1	p-hydroxybenzaldehyde	O=Cc1ccc(cc1)O
ethyl formate	CCOC=O	salicylaldehyde	c1cc(c(cc1)O)C=O
methyl acetate	CC(=O)OC	benzyl chloride	ClCc1ccccc1
propionic acid	OC(=O)CC	benzyl alcoholb	OCc1ccccc1
1-bromopropane	CCCB	m-cresol	Cc1cc(O)ccc1
isopropyl chloride	ClC(C)C	o-cresol	Oc1c(cccc1)C
n-propyl chloride	C(CC)Cl	p-cresol	Cc1ccc(cc1)O
isopropyl iodide	C(C)(C)I	diethyl malonate	O(CC)C(=O)CC(=O)OCC
n-propyl iodide	CCCI	diisopropyl ketone	C(C(=O)C(C)C)(C)C
isopropanol	C(C)(C)O	1-heptanal	C(CCCC)CC=O
methyl ethyl ether	CCOC	2-heptanone	C(CCCC)C(=O)C
1-propanol	CCCO	4-heptanoneb	O=C(CCC)CCC
1,2-propyleneglycol	O[C@H](CO)C	1-methylcyclohexanol	OC1(CCCCC1)C
1,3-propyleneglycol	C(CO)CO	5-methyl-2-hexanone	C(CC(=O)C)C(C)C
isopropyl mercaptan	C(S)(C)C	n-butyl propionate	O(CCCC)C(=O)CC
n-propyl mercaptan	CCCS	ethyl isovalerate	C(C(=O)OCC)C(C)C
furan	c1ccccc1	n-heptanoic acid	C(CCCCC)C(=O)O
thiophene	s1ccccc1	n-hexyl formate	C(CCOCC=O)CCC
2,5-dihydrofuran	C1=CCOC1	n-propyl n-butyrate	O(C(=O)CCC)CCC
methacrylic acid	OC(=O)C(=C)C	1-heptanol	C(CCCC)CCO
1,2-dichlorobutane	C([C@H](Cl)CC)Cl	2-heptanol	O[C@@H](CCCC)C
2,3-dichlorobutane	Cl[C@@H]([C@H](Cl)C)C	5-methyl-1-hexanol	C(CCCO)C(C)C
1-butanalb	O=CCCC	n-heptyl mercaptan	CCCCCCCS
methyl ethyl ketone	CC(=O)CC	vanillin	COc1cc(C=O)ccc1O
2-methylpropanal	CC(C)C=O	p-ethylphenol	CCc1ccc(cc1)O
tetrahydrofuran	O1CCCC1	2,3-xyleneol	Oc1c(c(cc1)C)C
n-butyric acid	C(C(=O)O)CC	2,4-xyleneol	Cc1c(O)ccc(c1)C
ethyl acetate	O(CC)C(=O)C	2,5-xyleneol	Oc1c(ccc(c1)C)C
isobutyric acid	CC(C)C(=O)O	2,6-xyleneol	Oc1c(cccc1C)C
methyl propionate	O(C(=O)CC)C	3,4-xyleneol	Oc1cc(c(cc1)C)C
n-propyl formate	O(CCC)C=O	3,5-xyleneol	c1(C)cc(C)cc(c1)O
tetrahydrothiophene	C1CCCCS1	diethyl maleate	C(=C\C(=O)OCC)\C(=O)OCC
1-bromobutane	C(CCB)C	1-Octanal	C(CCCC=O)CCC
2-bromobutane	C([C@@H](Br)C)C	2-octanone	C(CCCCC)C(=O)C
n-butyl chloride	C(CCC)Cl	n-butyl n-butyrate	CCCC(=O)OCCCC
sec-butyl chloride	C[C@@H](Cl)CC	n-heptyl formate	O(CCCCCC)C=O
tert-butyl chloride	ClC(C)(C)C	n-hexyl acetate	CCCCCOC(=O)C

isobutyl chloride	<chem>CC(CCl)C</chem>	isobutyl isobutyrate	<chem>O(CC(C)C)C(=O)C(C)C</chem>
1-butanol	<chem>OCCCC</chem>	di-sec-butyl ether	<chem>O([C@@H](CC)C)[C@@H](CC)C</chem>
2-butanol	<chem>O[C@H](CC)C</chem>	2-ethyl-1-hexanol	<chem>C([C@H](CC)CO)CCC</chem>
diethyl etheF	<chem>O(CC)CC</chem>	1-octanol	<chem>C(CCCCO)CCC</chem>
2-methyl-1-propanol	<chem>C(CO)(C)C</chem>	2-octanol	<chem>C(CCCCC)[C@H](C)O</chem>
2-methyl-2-propanol	<chem>C(C)(C)(C)O</chem>	n-octyl mercaptan	<chem>C(CCCCS)CCC</chem>
methyl n-propyl ether	<chem>COCCC</chem>	ethyl benzoate	<chem>O(CC)C(=O)c1ccccc1</chem>
1,3-butanediol	<chem>O[C@H](C)CCO</chem>	benzyl ethyl ether	<chem>O(Cc1ccccc1)CC</chem>
1,4-butanediol	<chem>OCCCCO</chem>	isophorone	<chem>CC1=CC(=O)CC(Cl)(C)C</chem>
2,3-butanediol	<chem>[C@H]([C@@H](C)O)(C)O</chem>	diisobutyl ketone	<chem>C(C(=O)CC(C)C)C(C)C</chem>
n-butyl mercaptan	<chem>C(CCS)C</chem>	2-nonanone	<chem>CCCCCCCC(=O)C</chem>
sec-butyl mercaptana	<chem>S[C@@H](CC)C</chem>	5-nonanone	<chem>O=C(CCCC)CCCC</chem>
tert-butyl mercaptan	<chem>C(C)(C)(C)S</chem>	n-heptyl acetate	<chem>C(CCCCC)COC(=O)C</chem>
isobutyl mercaptan	<chem>CC(C)CS</chem>	n-nonanoic acid	<chem>CCCCCCCCC(=O)O</chem>
methyl n-propyl sulfide	<chem>S(CCC)C</chem>	n-octyl formate	<chem>O(CCCCCCCC)C=O</chem>
2-methylthiophene	<chem>c1cc(s1)C</chem>	2,6-dimethyl-4-heptanol	<chem>C(C(CC(C)C)O)C(C)C</chem>
methyl isopropyl ketone	<chem>O=C(C)C(C)C</chem>	1-nonanol	<chem>C(CCCCC)CCCO</chem>
1-pentanal	<chem>C(CCC)C=O</chem>	2-nonanol	<chem>O[C@@H](CCCCCCC)C</chem>
2-pentanone	<chem>CCCC(=O)C</chem>	n-nonyl mercaptan	<chem>C(CCCCC)CCCS</chem>
3-pentanone	<chem>O=C(CC)CC</chem>	dimethyl phthalate	<chem>O(C(=O)c1c(C(=O)OC)cccc1)C</chem>
n-butyl formate	<chem>O(CCCC)C=O</chem>	anethole	<chem>c1cc(ccc1/C=C/C)OC</chem>
sec-butyl formate	<chem>O([C@@H](CC)C)C=O</chem>	p-tert-butylphenol	<chem>Oc1ccc(C(C)(C)C)cc1</chem>
tert-butyl formate	<chem>O(C(C)(C)C)C=O</chem>	n-decanoic acid	<chem>C(CCCCCC)CCC(=O)O</chem>
ethyl propionate	<chem>O(C(=O)CC)CC</chem>	isopentyl isovalerate	<chem>O(CCC(C)C)C(=O)CC(C)C</chem>
isobutyl formate	<chem>O(CC(C)C)C=O</chem>	n-octyl acetate	<chem>C(CCCCCC)COC(=O)C</chem>
isopropyl acetate	<chem>O(C(C)C)C(=O)C</chem>	1-decanol	<chem>C(CCCCCO)CCCC</chem>
methyl n-butyrate	<chem>C(C(=O)OC)CC</chem>	n-decyl mercaptan	<chem>C(CCCCCS)CCCC</chem>
isovaleric acid	<chem>C(C(C)C)C(=O)O</chem>	2-ethylhexyl acrylate	<chem>C([C@H](CCCC)CC)OC(=O)C=C</chem>
neopentanoic acid	<chem>OC(=O)C(C)(C)C</chem>	n-undecanoic acid	<chem>C(CCCCC(=O)O)CCCCC</chem>
n-pentanoic acid	<chem>CCCCC(=O)O</chem>	methyl decanoate	<chem>CCCCCCCCC(=O)OC</chem>
n-propyl acetate	<chem>O(C(=O)C)CCC</chem>	1-undecanol	<chem>C(CCCCCC)CCCO</chem>
1-chloropentane	<chem>C(Cl)CCCC</chem>	diethyl phthalate	<chem>O(C(=O)c1c(ccc1)C(=O)OCC)CC</chem>
2,2-dimethyl-1-propanol	<chem>OCC(C)(C)C</chem>	n-decyl acetate	<chem>C(CCCCCCCC)COC(=O)C</chem>
ethyl isopropyl ether	<chem>O(C(C)C)CC</chem>	n-dodecanoic acid	<chem>C(CCCCCCCC)CCC(=O)O</chem>
ethyl propyl ether	<chem>CCOCCC</chem>	di-n-hexyl ether	<chem>C(OCCCCC)CCCCC</chem>
2-methyl-1-butanol	<chem>C[C@H](CO)CC</chem>	n-tridecanoic acid	<chem>OC(=O)CCCCCCCCCCCC</chem>
2-methyl-2-butanol	<chem>OC(CC)(C)C</chem>	anthraquinone	<chem>O=C1c2c(C(=O)c3c1cccc3)cccc2</chem>
3-methyl-1-butanol	<chem>CC(C)CCO</chem>	p-tert-octylphenol	<chem>Oc1ccc(C(CC(C)(C)C)(C)C)cc1</chem>

Cuadro 7.2 – Nombres y SMILES de las moléculas de la base de datos de 276 compuestos.

La tercer base de datos contuvo originalmente 496 especies químicas, de las cuales su solubilidad acuosa es conocida. Esta puede encontrarse en la referencia [56]. En distintas etapas, varias especies fueron removidas de la base de datos original.

La D-fructosa (CAS:57-48-7) fue removida, ya que en solución acuosa se encuentra mayormente en dos formas disferentes: 70 % como fructopiranososa y 22 % como fructofuranososa. La correcta predicción del espectro requeriría un promediado de las diferentes formas en las que esta especie se encuentra. Dado que es estadísticamente poco relevante, se prefirió remover a esta especie.

A todas las especies restantes se les optimizó la geometría y se les realizó un estudio conformacional. Durante el cálculo de las propiedades magnéticas se produjeron fallos en varias moléculas¹. Estos cálculos resultaban interrumpidos súbitamente sin mensaje de error alguno, en ocasiones luego de días de haber comenzado los respectivos cálculo. No se encontró la causa, pero se verificó que no era falta de memoria RAM ni espacio para escritura en disco. Esto no fue considerado como un problema de importancia, dada la gran cantidad de especies químicas en la base de datos. Se conservaron entonces 464 especies, cuyos nombres están tabulados junto a sus correspondientes SMILES en la Tabla

¹Las moléculas problemáticas fueron, según el orden de las moléculas en la base de datos del artículo original, las siguientes: 78, 80, 82, 113, 205, 210, 214, 308, 340, 341, 347, 350, 351, 360, 365, 394, 396, 399, 400, 421, 444, 460, 469, 479, 480, 486, 488, 489.

hexane	CCCCC
3-methylpentane	C(C)(CC)CC
cyclohexane	C1CCCCC1
2,4-dimethylpentane	C(C(C)C)C(C)C
octane	C(CCCC)CCC
cyclooctane	C1CCCCCCC1
pent-2-ene(e)-pent-2-ene	C(/C=C/C)C
4-methylpent-1-ene	C(C(C)C)C=C
non-1-ene	C(CCCC=C)CCC
penta-1,4-diene	C(C=C)C=C
hexa-1,5-diene	C(CC=C)C=C
buta-1,3-diene	C(=C)C=C
cyclohexa-1,4-diene	C1C=CCC=C1
1-octyne	C(CCCC)CC#C
1-nonyne	C(CCCC#C)CCC
1,4-dimethylbenzene	Cc1ccc(cc1)C
1,3-dimethylbenzene	Cc1cc(C)ccc1
1,3,5-trimethylbenzene	c1(C)cc(C)cc(c1)C
1,2,3-trimethylbenzene	Cc1c(C)cccc1C
1-ethyl-4-methylbenzene	CCc1ccc(cc1)C
(propan-2-yl)benzene	CC(C)c1ccccc1
butylbenzene	C(Cc1ccccc1)CC
1,4-diethylbenzene	CCc1ccc(CC)cc1
(r)-sec-butylbenzene	C([C@@H](C)c1ccccc1)C
1,2,3,4,5-pentamethylbenzene	Cc1c(C)c(C)c(cc1C)C
2-methylbutan-2-ylbenzene	C(C(C)(C)c1ccccc1)C
hexylbenzene	C(CCCC)Cc1ccccc1
ethenylbenzene	c1(ccccc1)C=C
2-phenylethylbenzene	C(Cc1ccccc1)c1ccccc1
2-methylnaphthalene	c1(C)cc2c(cc1)cccc2
1-ethylnaphthalene	C(C)c1cccc2cccc12
2-ethylnaphthalene	CCc1cc2c(cc1)cccc2
1,4-dimethylnaphthalene	c1(C)c2c(c(cc1)C)cccc2
2,6-dimethylnaphthalene	c1(C)cc2c(cc1)cc(cc2)C
1,4,5-trimethylnaphthalene	c12c(c(ccc1C)C)cccc2C
phenanthrene	c1cc2c(cc1)ccc1c2cccc1
9-methylanthracene	Cc1c2c(cc3c1cccc3)cccc2
fluoranthene	c1ccc2c3c1c1c(c3ccc2)cccc1
11h-benzo[b]fluorene	c12c3c(Cc1cc1c(c2)cccc1)cccc3
triphenylene	c12c3c(c4c(c1cccc2)cccc4)cccc3
benzo[a]pyrene	c1ccc2c(c1)cc1c3c2ccc2c3c(cc1)ccc2
7,12-dimethylbenzo[b]phenanthrene	c12c(cccc1)c(c1c(c2C)ccc2c1cccc2)C
benzo[e]pyrene	c12c3c4c5c(c1cccc2ccc3ccc4)cccc5
chrysene	c12c(c3c(cc2)cccc3)ccc2c1cccc2
perylene	c12c3c4c5c(c1cccc2ccc3)cccc5ccc4
1,12-benzoperylene	c12c3c(ccc1)ccc1c3c3c4c2cccc4ccc3cc1
anthracen-2-amine	Nc1cc2c(cc3c(c2)cccc3)cc1
2-ethylanthracene	C(C)c1cc2c(cc1)cc1c(c2)cccc1
benz[j]fluoranthene	c12c3c4c(c1ccc1c2cccc1)cccc4ccc3
benzo[k]fluoranthene	c1ccc2c3c1c1c(c3ccc2)cc2c(c1)cccc2
bromo(chloro)methane	BrCCl
dibromomethane	BrCBr
chloroform	C(Cl)(Cl)Cl
dibromo-chloromethane	BrC(Br)Cl
tetrachloromethane	C(Cl)(Cl)(Cl)Cl
tetrabromomethane	BrC(Br)(Br)Br
1,1-dichloroethane	C(C)(Cl)Cl
1-bromo-2-chloroethane	ClCCBr
1,1,2-trichloroethane	ClC(Cl)CCl
1,1,2,2-tetrachloroethane	ClC(Cl)C(Cl)Cl
1,1,1,2,2,2-hexachloroethane	ClC(Cl)(Cl)C(Cl)(Cl)Cl
1-bromopropane	CCCB
1,3-dibromopropane	C(CCB)Br
(2r)-1,2-dibromo-3-chloro-propane	C([C@@H](CCl)Br)Br

(2r)-2-chlorobutane	C[C@@H](Cl)CC
1-bromobutane	C(CCCBr)C
1-bromo-2-methylpropane	BrCC(C)C
1-chloropentane	C(Cl)CCCC
3-chloropentane	ClC(CC)CC
2-chloro-2-methylbutane	ClC(CC)(C)C
1-bromopentane	CCCCCBr
1-bromooctane	CCCCCCCCBr
1,2-dichloro-1,1,2,2-tetrafluoroethane	ClC(F)(F)C(Cl)(F)F
(1r,2s,3r,4r,5s,6s)-1,2,3,4,5,6-hexachlorocyclohexane	[C@@H]1([C@H]([C@@H]([C@H]([C@@H]([C@H]1Cl)Cl)Cl)Cl)Cl)Cl
1-bromoheptane	C(CCCCCBr)C
1-chloroheptane	ClCCCCCCC
1,1,2,2-tetrachloroethene	C(=C(Cl)Cl)(Cl)Cl
aldrin	[C@H]12[C@H]3[C@@H]([C@H](C=C1)C2)[C@@]1(C(=C([C@]3(Cl(Cl)Cl)Cl)Cl)Cl)Cl
3-bromoprop-1-ene	BrCC=C
chlorobenzene	Clc1ccccc1
1-bromo-3-chlorobenzene	c1c(cc(cc1)Cl)Br
1-bromo-4-chlorobenzene	c1c(ccc(c1)Br)Cl
1,2-dibromobenzene	BrC1c(Br)cccc1
1,2,3-trichlorobenzene	c1cc(c(c(c1)Cl)Cl)Cl
1,3,5-trichlorobenzene	c1c(cc(cc1Cl)Cl)Cl
1,2,3-tribromobenzene	BrC1c(Br)cccc1Br
1,2,4,5-tetrafluorobenzene	c1(cc(c(cc1F)F)F)F
1,2,3,4-tetrachlorobenzene	Clc1c(Cl)c(Cl)ccc1Cl
1,2,3,4,5-pentachlorobenzene	Clc1c(Cl)c(Cl)cc(Cl)c1Cl
1-bromo-4-methylbenzene	BrC1ccc(cc1)C
1,4-difluorobenzene	Fc1ccc(F)cc1
chloromethylbenzene	ClCc1ccccc1
1-chloro-4-methylbenzene	Clc1ccc(cc1)C
1-chloro-2-phenylbenzene	c1cc(ccc1)c1ccccc1Cl
1-chloro-3-(3-chlorophenyl)benzene	c1(cc(ccc1)c1ccc(c1)Cl)Cl
1,2-dichloro-4-phenylbenzene	c1cc(ccc1)c1ccc(c(c1)Cl)Cl
1,2,4-trichloro-5-phenylbenzene	c1cc(ccc1)c1cc(c(cc1Cl)Cl)Cl
1,3,5-trichloro-2-phenylbenzene	c1cc(ccc1)c1c(cc(cc1Cl)Cl)Cl
1,4-dichloro-2-(3-chlorophenyl)benzene	c1(cc(ccc1)c1cc(ccc1Cl)Cl)Cl
1,2-dichloro-4-(2-chlorophenyl)benzene	c1c(c(ccc1)c1ccc(c(c1)Cl)Cl)Cl
2,4-dichloro-1-(4-chlorophenyl)benzene	c1cc(ccc1Cl)c1ccc(cc1Cl)Cl
1,2-dichloro-3-(2,5-dichlorophenyl)benzene	c1c(c(cc(c1)Cl)c1cccc(c1Cl)Cl)Cl
1,4-dichloro-2-(2,5-dichlorophenyl)benzene	c1c(c(cc(c1)Cl)c1cc(c(cc1Cl)Cl)Cl)Cl
1,4-dichloro-2-(2,6-dichlorophenyl)benzene	c1c(c(c(cc1)Cl)c1cc(c(cc1Cl)Cl)Cl)Cl
1,3-dichloro-2-(2,6-dichlorophenyl)benzene	c1c(c(c(cc1)Cl)c1c(ccc1Cl)Cl)Cl
1,2,3-trichloro-4-(2,3-dichlorophenyl)benzene	c1(c(c(ccc1)c1ccc(c(c1Cl)Cl)Cl)Cl)Cl
1,2,4-trichloro-5-(3,4-dichlorophenyl)benzene	c1(cc(ccc1Cl)c1cc(c(cc1Cl)Cl)Cl)Cl
1,2,3,4,5-pentachloro-6-phenylbenzene	c1cc(ccc1)c1c(c(c(c(c1Cl)Cl)Cl)Cl)Cl
1,2,3-trichloro-4-(2,4,5-trichlorophenyl)benzene	c1c(c(cc(c1Cl)Cl)c1ccc(c(c1Cl)Cl)Cl)Cl
1,2,3,4-tetrachloro-5-(2,5-dichlorophenyl)benzene	c1c(c(cc(c1)Cl)c1cc(c(c(c1Cl)Cl)Cl)Cl)Cl
1,3,5-trichloro-2-(2,4,6-trichlorophenyl)benzene	c1c(c(c(cc1Cl)Cl)c1c(cc(cc1Cl)Cl)Cl)Cl
1,2,3,5-tetrachloro-4-(3,4-dichlorophenyl)benzene	c1(cc(ccc1Cl)c1c(cc(c(c1Cl)Cl)Cl)Cl)Cl
1,2,3-trichloro-4-(2,3,4-trichlorophenyl)benzene	c1(c(c(ccc1Cl)c1ccc(c(c1Cl)Cl)Cl)Cl)Cl
1,2,3,5-tetrachloro-4-(2,4,5-trichlorophenyl)benzene	c1c(c(cc(c1Cl)Cl)c1c(cc(c(c1Cl)Cl)Cl)Cl)Cl
1,2,3,4,5-pentachloro-6-(2,5-dichlorophenyl)benzene	c1c(c(cc(c1)Cl)c1c(c(c(c1Cl)Cl)Cl)Cl)Cl
1,2,3,4-tetrachloro-5-(2,3,4,5-tetrachlorophenyl)benzene	c1(c(c(cc(c1Cl)Cl)c1cc(c(c(c1Cl)Cl)Cl)Cl)Cl)Cl
1,2,3,4,5-pentachloro-6-(2,3,4,5-tetrachlorophenyl)benzene	c1(c(c(cc(c1Cl)Cl)c1c(c(c(c1Cl)Cl)Cl)Cl)Cl)Cl
1-chloro-4-[2,2-dichloro-1-(4-chlorophenyl)ethyl]benzene	ClC(Cl)C(Clccc(Cl)cc1)c1ccc(Cl)cc1
pentan-3-ol	OC(CC)CC
2-methylbutan-2-ol	OC(CC)(C)C
(2s)-3-methyl-2-butanol	CC([C@H](C)O)C
(2r)-2-hexanol	O[C@@H](CCCC)C
2,3-dimethyl-1-butanol	CC([C@H](C)CO)C
(3r)-2-methyl-3-pentanol	O[C@@H](C(C)C)CC
3,3-dimethylbutan-1-ol	OCCC(C)(C)C
(2r)-3,3-dimethyl-2-butanol	[C@H](C(C)(C)C)(C)O
3-methylpentan-3-ol	OC(CC)(CC)C
heptan-1-ol	C(CCCC)CO

(3r)-3-heptanol	<chem>O[C@@H](CCCC)CC</chem>
(2r)-2,4-dimethyl-1-pentanol	<chem>OC[C@H](CC(C)C)C</chem>
2-methylhexan-2-ol	<chem>OC(CCCC)(C)C</chem>
2,4-dimethylpentan-3-ol	<chem>OC(C(C)C)C(C)C</chem>
2,3-dimethyl-2-pentanol	<chem>[C@H](C(C)(C)O)(CC)C</chem>
2,3,3-trimethylbutan-2-ol	<chem>OC(C(C)(C)C)(C)C</chem>
(3r)-3-methyl-3-heptanol	<chem>O[C@@](CCCC)(CC)C</chem>
(3s)-2,2,3-trimethyl-3-pentanol	<chem>CC[C@@](O)(C)C(C)(C)C</chem>
2-methylheptan-2-ol	<chem>C(C(C)(C)O)CCCC</chem>
dodecan-1-ol	<chem>C(CCCCCCO)CCCCC</chem>
3-phenylpropan-1-ol	<chem>C(Cc1cccc1)CO</chem>
nonan-1-ol	<chem>C(CCCCC)CCCCO</chem>
trans-4-methylcyclohexanol	<chem>O[C@H]1CC[C@@H](C1)C</chem>
benzene-1,3-diol	<chem>Oc1cc(O)ccc1</chem>
benzene-1,4-diol	<chem>Oc1ccc(cc1)O</chem>
2-methylphenol	<chem>Oc1c(ccc1)C</chem>
phenylmethanol	<chem>OCc1ccccc1</chem>
3-methylphenol	<chem>Cc1cc(O)ccc1</chem>
2,4-dimethylphenol	<chem>Cc1c(O)ccc(c1)C</chem>
3,5-dimethylphenol	<chem>c1(C)cc(C)cc(c1)O</chem>
4-tert-butylphenol	<chem>Oc1ccc(C(C)(C)C)cc1</chem>
naphthalen-1-ol	<chem>Oc1c2c(ccc1)cccc2</chem>
naphthalene-1,5-diol	<chem>Oc1c2c(c(O)ccc2)ccc1</chem>
2-phenylphenol	<chem>Oc1c(c2ccccc2)cccc1</chem>
hexanal	<chem>O=CCCCC</chem>
nonanal	<chem>C(CCCCC)CCC=O</chem>
benzaldehyde	<chem>O=Cc1ccccc1</chem>
cyclohexanone	<chem>O=C1CCCCC1</chem>
heptan-2-one	<chem>C(CCCC)C(=O)C</chem>
nonan-5-one	<chem>O=C(CCCC)CCCC</chem>
5-methylhexan-2-one	<chem>C(CC(=O)C)C(C)C</chem>
cyclohexa-2,5-diene-1,4-dione	<chem>O=C1C=CC(=O)C=C1</chem>
pentane-2,4-dione	<chem>C(C(=O)C)C(=O)C</chem>
2-decanone	<chem>O=C(CCCCCCCC)C</chem>
4-pregnen-3,20-dione	<chem>O=C([C@@H]1[C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)CC4)CC3)C)CC2)CC1)C)C</chem>
anthracene-9,10-dione	<chem>O=C1c2c(C(=O)c3cccc3)cccc2</chem>
pentanedioic acid	<chem>C(CCC(=O)O)C(=O)O</chem>
hexanedioic acid	<chem>C(CCC(=O)O)CC(=O)O</chem>
octanoic acid	<chem>C(CCC(=O)O)CCCC</chem>
decanoic acid	<chem>C(CCCCC)CCC(=O)O</chem>
benzoic acid	<chem>OC(=O)c1ccccc1</chem>
phthalic acid	<chem>OC(=O)c1c(ccc1)C(=O)O</chem>
hexadecanoic acid	<chem>OC(=O)CCCCCCCCCCCCCCC</chem>
(+)-gibberellic acid	<chem>O1[C@]23[C@H]4[C@@]5([C@H]([C@@H]2[C@@]([C@@H](O)C=C3)(C1=O)C)C(=O)O)C[C@@](O)(CC4)C(=C)C5</chem>
2-ethylbutanoic acid	<chem>OC(=O)C(CC)CC</chem>
2-propylpentanoic acid	<chem>OC(=O)C(CCC)CCC</chem>
2-(phenoxy)acetic acid	<chem>C(Oc1ccccc1)C(=O)O</chem>
undecanoic acid	<chem>C(CCCCC(=O)O)CCCCC</chem>
tetradecanoic acid	<chem>OC(=O)CCCCCCCCCCCCCCC</chem>
methyl formate	<chem>COC=O</chem>
methyl butanoate	<chem>C(C(=O)OC)CC</chem>
2-methylpropyl formate	<chem>O(CC(C)C)C=O</chem>
butyl acetate	<chem>O(CCCC)C(=O)C</chem>
methyl pentanoate	<chem>O(C(=O)CCCC)C</chem>
2-methylpropyl prop-2-enoate	<chem>O(C(=O)C=C)CC(C)C</chem>
propyl butanoate	<chem>O(C(=O)CCC)CCC</chem>
hexyl acetate	<chem>CCCCCOC(=O)C</chem>
ethyl hexanoate	<chem>CCCCC(=O)OCC</chem>
ethyl heptanoate	<chem>C(C(=O)OCC)CCCCC</chem>
ethyl octanoate	<chem>C(C(=O)OCC)CCCCC</chem>
ethyl nonanoate	<chem>C(C(=O)OCC)CCCCCCC</chem>
ethyl decanoate	<chem>C(C(CCCCCC)C(=O)OCC</chem>
methyl benzoate	<chem>COC(=O)c1ccccc1</chem>

bis(2-methylpropyl) benzene-1,2-dicarboxylate	<chem>O(CC(C)C)C(=O)c1c(C(=O)OCC(C)C)cccc1</chem>
2-methoxypropane	<chem>COC(C)C</chem>
2-methoxy-2-methylpropane	<chem>O(C(C)(C)C)C</chem>
1-propoxypropane	<chem>C(OCCC)CC</chem>
2-propan-2-yloxypropane	<chem>O(C(C)C)C(C)C</chem>
1,2-diethoxyethane	<chem>O(CCOCC)CC</chem>
phenoxybenzene	<chem>O(c1ccccc1)c1ccccc1</chem>
oxanthrene	<chem>c12c(Oc3c(O1)cccc3)cccc2</chem>
oxolane	<chem>O1CCCC1</chem>
cortisol	<chem>O[C@]1([C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)CC4)CC3)C)[C@@H](O)C2)CC1)C(=O)CO</chem>
dexamethasone	<chem>F[C@@]12[C@H]([C@H]3[C@@]([C@](O)([C@@H](C3)C)C(=O)CO)(C[C@@H]1O)C)CCC1=CC(=O)C=C[C@]21C</chem>
(11)-11,17-dihydroxy-3,20-dioxopregn-4-en-21-yl acetate	<chem>O[C@]1([C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)CC4)CC3)C)[C@@H](O)C2)CC1)C(=O)COC(=O)C</chem>
prednisolone	<chem>O[C@]1([C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)C=C2)CC3)C)CC1)C)C(=O)CO</chem>
3-hydroxyestra-1(10),2,4-trien-17-one	<chem>O=C1[C@@]2([C@H]([C@H]3[C@H](CC2)c2c(CC3)cc(O)cc2)CC1)C</chem>
androstenedione	<chem>O=C1[C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)CC4)CC3)C)CC2)CC1)C</chem>
11-deoxycorticosterone	<chem>O=C([C@H]1[C@@]2([C@H]([C@H]3[C@@H]([C@@]4(C(=CC(=O)CC4)CC3)C)CC2)CC1)C)CO</chem>
fludrocortisone	<chem>F[C@]12[C@H]([C@H]3[C@@]([C@](O)(CC3)C(=O)CO)(C[C@@H]2O)C)CCC2=CC(=O)CC[C@]12C</chem>
fluorometholone	<chem>F[C@@]12[C@H]([C@H]3[C@@]([C@](O)(CC3)C(=O)C)(C[C@@H]1O)C)C[C@@H](C1=CC(=O)C=C[C@]21C)C</chem>
betamethasone 17-valerate	<chem>F[C@@]12[C@H]([C@H]3[C@@]([C@](OC(=O)CCCC)([C@H](C3)C)C(=O)CO)(C[C@@H]1O)C)CCC1=CC(=O)C=C[C@]21C</chem>
2-methoxyphenol	<chem>COc1ccccc1O</chem>
4-methoxybenzaldehyde	<chem>O=Cc1ccc(cc1)OC</chem>
4-chlorophenol	<chem>c1c(ccc(c1)Cl)O</chem>
2,4,6-trichlorophenol	<chem>c1c(c(c(cc1Cl)Cl)O)Cl</chem>
4-chloro-3-methylphenol	<chem>Clc1c(cc(O)cc1)C</chem>
1-chloro-2-(2-chloroethoxy)ethane	<chem>ClCCOCCCl</chem>
dieldrin	<chem>[C@H]12[C@H]3[C@@H]([C@H](Cl)[C@@H]1[C@H]2O1)[C@@]([C@@]1([C@]3(Cl(Cl)Cl)Cl)Cl)([C]1)Cl</chem>
hexan-1-amine	<chem>C(CCCN)CC</chem>
n-butylbutan-1-amine	<chem>CCCCNCCCC</chem>
octan-1-amine	<chem>C(CCCCN)CCC</chem>
benzene-1,4-diamine	<chem>Nc1ccc(N)cc1</chem>
2-methylaniline	<chem>Nc1c(ccc1)C</chem>
n-methylaniline	<chem>N(c1ccccc1)C</chem>
3-methylaniline	<chem>Nc1cc(ccc1)C</chem>
n,n-diethylaniline	<chem>N(CC)(CC)c1ccccc1</chem>
4-[(4-aminophenyl)methyl]aniline	<chem>Nc1ccc(Cc2ccc(N)cc2)cc1</chem>
pyridine	<chem>n1ccccc1</chem>
2,5-dimethylpyridine	<chem>n1c(ccc(c1)C)C</chem>
2,6-dimethylpyridine	<chem>n1c(cccc1C)C</chem>
3,4-dimethylpyridine	<chem>n1cc(c(cc1)C)C</chem>
3,5-dimethylpyridine	<chem>n1cc(cc(C)c1)C</chem>
4-pyridin-4-ylpyridine	<chem>n1ccc(c2ccncc2)cc1</chem>
pyridine-4-carbohydrazide	<chem>O=C(NN)c1ccncc1</chem>
benzo[f]quinoline	<chem>n1c2c(c3c(cc2)cccc3)ccc1</chem>
1,2,3-benzotriazole	<chem>c1ccc2n[nH]nc2c1</chem>
pyrrole	<chem>[nH]1cccc1</chem>
thiophene	<chem>s1cccc1</chem>
indole	<chem>c12c([nH]cc1)cccc2</chem>
pyridazine	<chem>c1ccnnc1</chem>
5-quinolinol	<chem>c1(O)c2c(ccc1)nc2</chem>
quinolin-6-ol	<chem>Oc1cc2c(nc2)cc1</chem>
5,5-diethyl-1-methyl-1,3-diazinane-2,4,6-trione	<chem>O=C1N(C(=O)NC(=O)Cl(CC)CC)C</chem>
4-butyl-1,2-di(phenyl)pyrazolidine-3,5-dione	<chem>O=C1N(N(C(=O)C1CCCC)c1ccccc1)c1ccccc1</chem>
3-(2-chlorophenothiazin-10-yl)-n,n-dimethylpropan-1-amine	<chem>Clc1cc2N(CCCN(C)C)c3c(Sc2cc1)cccc3</chem>
(3-exo)-8-methyl-8-azabicyclo[3.2.1]oct-3-yl (2s)-3-hydroxy-2-phenylpropanoate	<chem>O([C@@H]1C[C@@H]2N([C@@H](CC2)Cl)C(=O)[C@@H](c1ccccc1)CO</chem>

Bases de datos

4-chloro-2-(furan-2-ylmethylamino)-5-sulfamoylbenzoic acid	<chem>Clc1c(S(=O)(=O)N)cc(c(NCc2occc2)c1)C(=O)O</chem>
(E)-Diethylstilbestrol	<chem>Oc1ccc(/C(=C(CC)/c2ccc(O)cc2)/CC)cc1</chem>
5,5-diethyl-1,3-diazinane-2,4,6-trione	<chem>O=CINC(=O)NC(=O)Cl(CC)CC</chem>
[2-(carbamoyloxymethyl)-2-methylpentyl] carbamate	<chem>NC(=O)OCC(COC(=O)N)(C)CCC</chem>
4-amino-n-(4,6-dimethylpyrimidin-2-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1nc(cc(n1)C)C)c1ccc(N)cc1</chem>
2-[4-[3-(2-chlorophenothiazin-10-yl)propyl]piperazin-1-yl]ethanol	<chem>N1(CCCN2CCN(CC2)CCO)c2c(Sc3c1cc(Cl)cc3)cccc2</chem>
n,n-dimethyl-3-phenothiazin-10-ylpropan-1-amine	<chem>S1c2c(N(CCCN(C)C)c3c1cccc3)cccc2</chem>
6-chloro-1,1-dioxo-4h-benzo[e][1,2,4]thiadiazine-7-sulfonamide	<chem>Clc1c(S(=O)(=O)N)cc2S(=O)(=O)N=CNc2c1</chem>
n-(4-ethoxyphenyl)acetamide	<chem>CCOc1ccc(cc1)NC(=O)C</chem>
4-aminobenzenesulfonamide	<chem>S(=O)(=O)(N)c1ccc(N)cc1</chem>
tubercidin	<chem>O1[C@@H](n2c3ncnc(N)c3cc2)[C@H](O)[C@H](O)[C@H]1CO</chem>
4-amino-n-(1,3-thiazol-2-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1scn1)c1ccc(N)cc1</chem>
5-propan-2-yl-5-prop-2-enyl-1,3-diazinane-2,4,6-trione	<chem>O=CINC(=O)NC(=O)Cl(C(C)C)CC=C</chem>
5-allyl-5-[(2r)-2-pentanyl]-2-thioxodihydro-4,6(1h,5h)-pyrimidinedione	<chem>S=CINC(=O)C([C@@H](CCC)C)(CC=C)C(=O)N1</chem>
5-butyl-5-ethyl-1,3-diazinane-2,4,6-trione	<chem>O=CINC(=O)NC(=O)Cl(CCCC)CC</chem>
4-amino-n-(6-methoxypyridazin-3-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1nncc(OC)cc1)c1ccc(N)cc1</chem>
chromen-2-one	<chem>O=c1ccc2ccccc2o1</chem>
4-amino-n-(5-ethyl-1,3,4-thiadiazol-2-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1sc(nn1)CC)c1ccc(N)cc1</chem>
5-phenyl-5-prop-2-enyl-1,3-diazinane-2,4,6-trione	<chem>C(C=C)Cl(C(=O)NC(=O)NC1=O)c1cccc1</chem>
5-ethyl-5-phenyl-1,3-diazinane-4,6-dione	<chem>C1NC(=O)C(C(=O)N1)(CC)c1cccc1</chem>
Trichlormethiazide	<chem>Cl(Cl)c(S(=O)(=O)N)cc2c(Cl)N[C@@H](NS2(=O)=O)C(Cl)Cl</chem>
n-(4-aminophenyl)sulfonylacetamide	<chem>Cl(ccc(cc1)N)S(=O)(=O)NC(=O)C</chem>
meso-erythritol	<chem>OC[C@H](O)[C@@H](O)CO</chem>
allopurinol	<chem>Cl[nH]c(=O)c2c(n1n[nH]c2)O1[C@@H]2[C@H]3[C@@H]4N(c5c([C@@]64[C@H]4N(CC6)CC(=CC1)[C@@H]3C4)cc(OC)c(OC)c5)C(=O)C2</chem>
(-)-brucine	
4-amino-5-chloro-n-(2-diethylaminoethyl)-2-methoxybenzamide	<chem>Clc1cc(C(=O)NCCN(CC)CC)c(OC)cc1N</chem>
redupresin	<chem>S(=O)(=O)(N)c1sc2c(n1)ccc(OCC)c2</chem>
5-(1-cycloheptenyl)-5-ethyl-1,3-diazinane-2,4,6-trione	<chem>O=CINC(=O)NC(=O)Cl(C1=CCCCC1)CC</chem>
eriodictyol	<chem>O1[C@@H](CC(=O)c2c1cc(O)cc2O)c1cc(O)c(O)cc1</chem>
4-amino-n-(5-methoxypyrimidin-2-yl)benzenesulfonamide	<chem>c1cc(ccc1N)S(=O)(=O)Nc1ncc(cn1)OC</chem>
4-amino-n-(5-methyl-1,2-oxazol-3-yl)benzenesulfonamide	<chem>Cl(ccc(cc1)N)S(=O)(=O)Nc1noc(c1)C</chem>
5-[(3,4,5-trimethoxyphenyl)methyl]pyrimidine-2,4-diamine	<chem>O(c1cc(Cc2c(nc(nc2)N)N)cc(OC)c1OC)C</chem>
4-amino-n-(6-methoxy-2-methylpyrimidin-4-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1nc(nc(OC)c1)C)c1ccc(N)cc1</chem>
(r)-(-)-ibuprofen	<chem>Clc(ccc(c1)CC(C)C)[C@@H](C)C(=O)O</chem>
1,3,7-trimethylpurine-2,6-dione	<chem>O=c1n(c(=O)n(c2ncn(c12)C)C)C</chem>
quinolone	<chem>c12c(ccc1)[nH]c(=O)cc2</chem>
thymine	<chem>O=c1[nH]c(=O)[nH]cc1C</chem>
4-methylmorpholine	<chem>O1CCN(CC1)C</chem>
piperidine	<chem>C1CCNCC1</chem>
methyluracil	<chem>O=c1n(ccc(=O)[nH]1)C</chem>
1-methylpiperidine	<chem>N1(CCCCC1)C</chem>
2-aminophenol	<chem>Oc1c(N)cccc1</chem>
2-aminobenzoic acid	<chem>OC(=O)c1c(N)cccc1</chem>
ethyl carbamate	<chem>CCOC(=O)N</chem>
phenylmethyl carbamate	<chem>O(Cc1cccc1)C(=O)N</chem>
urea	<chem>C(=O)(N)N</chem>
methylurea	<chem>CNC(=O)N</chem>
1,1,3,3-tetramethylurea	<chem>O=C(N(C)C)N(C)C</chem>
phenylmethylurea	<chem>O=C(NC1cccc1)N</chem>
benzamide	<chem>O=C(N)c1cccc1</chem>
benzene-1,2-dicarboxamide	<chem>O=C(N)c1c(ccc1)C(=O)N</chem>
n-(4-hydroxyphenyl)acetamide	<chem>Oc1ccc(NC(=O)C)cc1</chem>
(2-propan-2-yloxyphenyl) n-methylcarbamate	<chem>O(C(C)C)c1c(OC(=O)NC)cccc1</chem>
(-)-codeine	<chem>O1[C@@H]2[C@H]34[C@H]([C@H](N(CC3)C)Cc3c4c1c(OC)cc3)C=C[C@@H]2O</chem>
1-nitropropane	<chem>CCCN(=O)=O</chem>
1-methyl-3-nitrobenzene	<chem>O=N(=O)c1cc(ccc1)C</chem>
n,n-dihydroxy-4-methylaniline	<chem>O=N(=O)c1ccc(cc1)C</chem>
4-nitrophenol	<chem>c1c(ccc(c1)N(=O)=O)O</chem>
2-(dihydroxyamino)phenol	<chem>c1(ccccc1O)N(=O)=O</chem>
3-nitrobenzoic acid	<chem>OC(=O)c1cc(N(=O)=O)ccc1</chem>
1-methoxy-2-nitrobenzene	<chem>O(c1c(N(=O)=O)cccc1)C</chem>
2-nitroaniline	<chem>O=N(=O)c1c(N)cccc1</chem>
4-nitroaniline	<chem>O=N(=O)c1ccc(N)cc1</chem>

4-chloroaniline	<chem>c1(ccc(cc1)Cl)N</chem>
3-chloroaniline	<chem>c1(cc(ccc1)Cl)N</chem>
6-chloro-n,n'-diethyl-1,3,5-triazine-2,4-diamine	<chem>c1(nc(nc(n1)NCC)NCC)Cl</chem>
propan-2-yl n-(3-chlorophenyl)carbamate	<chem>CC(C)OC(=O)Nc1cccc(Cl)c1</chem>
3-(4-chlorophenyl)-1-methoxy-1-methylurea	<chem>Clc1ccc(NC(=O)N(OC)C)cc1</chem>
n-(3,4-dichlorophenyl)propanamide	<chem>Clc1cc(NC(=O)CC)ccc1Cl</chem>
terbacil	<chem>Clc1c(=O)n(C(C)(C)C)c(=O)[nH]c1C</chem>
2,6-dinitro-n,n-dipropyl-4-(trifluoromethyl)aniline	<chem>FC(F)(F)c1cc(N(=O)=O)c(N(CCC)CCC)c(N(=O)=O)c1</chem>
(1r)-1-(4-chlorophenoxy)-3,3-dimethyl-1-(1h-1,2,4-triazol-1-yl)-2-butanone	<chem>Clc1ccc(O[C@@H](n2ncnc2)C(=O)C(C)(C)C)cc1</chem>
butane-1-thiol	<chem>SCCCC</chem>
benzenethiol	<chem>Sclccccc1</chem>
n'-ethyl-6-methylsulfanyl-n-propan-2-yl-1,3,5-triazine-2,4-diamine	<chem>S(c1nc(NC(C)C)nc(NCC)n1)C</chem>
n-tert-butyl-n'-ethyl-6-methylsulfanyl-1,3,5-triazine-2,4-diamine	<chem>S(c1nc(NC(C)(C)C)nc(NCC)n1)C</chem>
1,1-dioxo-1,2-benzothiazol-3-one	<chem>S1(=O)(=O)NC(=O)c2c1cccc2</chem>
2-methylbenzenesulfonamide	<chem>S(=O)(=O)(N)c1c(ccc1)C</chem>
phorate	<chem>S(P([S])(OCC)OCC)CSCC</chem>
prometryn	<chem>O=C1c2c(C(=O)N1CSP(OC)(OC)[S])cccc2</chem>
dimethoxy(4-nitrophenoxy)sulfanylphosphonium	<chem>[S]P(Oc1ccc(N(=O)=O)cc1)(OC)OC</chem>
dicaphton	<chem>Clc1c(OP([S])(OC)OC)ccc(N(=O)=O)c1</chem>
ethion	<chem>S(P([S])(OCC)OCC)CSP([S])(OCC)OCC</chem>
ronnel	<chem>Clc1c(OP([S])(OC)OC)cc(Cl)c(Cl)c1</chem>
hyodeoxycholic acid	<chem>O[C@@H]1[C@H]2[C@@]([C@@H]3[C@H]([C@H]4[C@@]([C@@H](CC4)[C@@H](CCC(=O)O)C)(CC3)C)Cl)(CC[C@H](O)C2)C</chem>
ursodeoxycholic acid	<chem>O[C@@H]1[C@H]2[C@H]3[C@@]([C@H](CC3)[C@@H](CCC(=O)O)C)(CC[C@H]2[C@@]([C@H](Cl)C[C@H](O)CC2)C)C</chem>
8-chloro-6-(2-chlorophenyl)-1-methyl-4h-[1,2,4]triazolo[4,3-a][1,4]benzodiazepine	<chem>Clc1cc2c(n3c(nnc3C)CN=C2c2c(Cl)cccc2)cc1</chem>
(r)-2-methylpiperazine	<chem>C[C@@H]1CNCCN1</chem>
n-(4-bromophenyl)acetamide	<chem>Brclccc(NC(=O)C)cc1</chem>
n-(4-nitrophenyl)acetamide	<chem>O=C(Nc1ccc(N(=O)=O)cc1)C</chem>
isonoruron	<chem>O=C(N[C@@H]1[C@@H]2[C@@H]([C@@H]3C[C@H]2CC3)CC1)N(C)C</chem>
4-chloro-2-[(5-chloro-2-hydroxyphenyl)methyl]phenol	<chem>Clc1cc(Cc2c(O)ccc(Cl)c2)c(O)cc1</chem>
3-amino-2,5-dichlorobenzoic acid	<chem>OC(=O)c1c(Cl)c(N)cc(Cl)c1</chem>
n,n-dimethyl-2,2-di(phenyl)acetamide	<chem>O=C(N(C)C)C(c1ccccc1)c1ccccc1</chem>
s-(2,3,3-trichloroprop-2-enyl)(di(propan-2-yl)amino)methanethioate	<chem>ClC(=C(Cl)Cl)CSC(=O)N(C(C)C)C(C)C</chem>
s-propyl (butyl-ethylamino)methanethioate	<chem>S(C(=O)N(CCCC)CC)CCC</chem>
butyl 4-hydroxybenzoate	<chem>O(CCCC)C(=O)c1ccc(O)cc1</chem>
methyl 4-hydroxybenzoate	<chem>O(C(=O)c1ccc(O)cc1)C</chem>
diethyl butanedioate	<chem>C(C(=O)OCC)CC(=O)OCC</chem>
butyl benzoate	<chem>O(C(=O)c1ccccc1)CCCC</chem>
methyl 4-aminobenzoate	<chem>O(C(=O)c1ccc(cc1)N)C</chem>
hexyl 4-aminobenzoate	<chem>O(CCCCCC)C(=O)c1ccc(N)cc1</chem>
octyl 4-aminobenzoate	<chem>O(CCCCCCCC)C(=O)c1ccc(N)cc1</chem>
n-(3-methylphenyl)acetamide	<chem>O=C(Nc1cc(ccc1)C)C</chem>
(2r)-2-ethylpiperidine	<chem>CC[C@H]1CCCCN1</chem>
1-methoxy-4-prop-2-enylbenzene	<chem>C(=C)Cc1ccc(cc1)OC</chem>
2,3,4,5-tetrachloro-6-methoxyphenol	<chem>Clc1c(OC)c(O)c(Cl)c(Cl)c1Cl</chem>
l-carvone	<chem>O=C1C[C@H](CC=C1C)C(=C)C</chem>
1,3-diazinane-2,4,5,6-tetrone	<chem>Cl(=O)NC(=O)C(=O)C(=O)N1</chem>
4-pyrimidinol	<chem>O=c1[nH]cncc1</chem>
pyrrolidine-2,5-dione	<chem>O=C1CCC(=O)N1</chem>
bis(2-chloroethyl) sulfoxide	<chem>ClCC[S]([O])CCCl</chem>
l-(-)-asparagine	<chem>NC(=O)C[C@H](N)C(=O)O</chem>
(s)-(+)-2-aminobutyric acid	<chem>OC(=O)[C@H](N)CC</chem>
pyridin-3-ol	<chem>c1ncc(cc1)O</chem>
4-pyridone	<chem>O=c1cc[nH]cc1</chem>
adenine	<chem>[nH]1c2c(nc1)ncnc2N</chem>
1,3-dichloro-5,5-dimethylimidazolidine-2,4-dione	<chem>ClN1C(C(=O)N(Cl)Cl=O)(C)C</chem>
6-methyl-2-thiouracil	<chem>S=c1[nH]c(cc(=O)[nH]1)C</chem>
5-ethyl-2,4-imidazolidinedione	<chem>O=C1NC(=O)N[C@@H]1CC</chem>
methazolamide	<chem>S(=O)(=O)(N)c1s/c(=NC(=O)C)/n(n1)C</chem>

Bases de datos

4,6,7-trichloro-2,1,3-benzothiadiazole	<chem>Clc1c2nsnc2c(Cl)cc1Cl</chem>
pteridine	<chem>n1c2c(nc1)ncnc2</chem>
2-pteridinethiol	<chem>S=c1[nH]c2nccnc2cn1</chem>
(3h)-pteridinethione	<chem>S=c1nc[nH]c2nccnc12</chem>
1-methyl-7,9-dihydro-1h-purine-2,6,8(3h)-trione	<chem>O=c1n(c(=O)[nH]c2[nH]c(=O)[nH]c12)C</chem>
benzenesulfonamide	<chem>S(=O)(=O)(N)c1ccccc1</chem>
isosorbide dinitrate	<chem>O1[C@H]2[C@H](OC[C@H]2ON(=O)=O)[C@@H](ON(=O)=O)C1</chem>
thiram	<chem>S(SC(=S)N(C)C)C(=S)N(C)C</chem>
inositol	<chem>O[C@@H]1[C@H](O)[C@H](O)[C@@H](O)[C@@H](O)[C@H]1O</chem>
pentyl carbamate	<chem>O(CCCCC)C(=O)N</chem>
2,2-dimethylpropyl carbamate	<chem>O(CCC(C)(C)C)C(=O)N</chem>
bromoxynil	<chem>C(#N)c1cc(Br)c(c1)BrO</chem>
4-methylpteridine	<chem>n1c2c(ncnc2ncc1)C</chem>
2-methoxypteridine	<chem>O(c1nc2nccnc2cn1)C</chem>
7-methoxypteridine	<chem>O(c1nc2nccnc2nc1)C</chem>
2-hydroxybenzaldehyde	<chem>c1cc(c(c1)O)C=O</chem>
3,4,5-trihydroxybenzoic acid	<chem>OC(=O)c1cc(O)c(c1)O</chem>
(3s)-3-methylcyclohexanone	<chem>C[C@@H]1CC(=O)CCC1</chem>
hexyl carbamate	<chem>O(CCCCCC)C(=O)N</chem>
2-(trifluoromethyl)benzoic acid	<chem>c1c(c(ccc1)C(=O)O)C(F)(F)F</chem>
2-(2-chlorophenoxy)acetic acid	<chem>Clc1c(OCC(=O)O)cccc1</chem>
cyclohexyl acetate	<chem>O(C1CCCCC1)C(=O)C</chem>
4-hydroxy-3-methoxybenzoic acid	<chem>COc1cc(C(=O)O)ccc1O</chem>
1-(4-aminophenyl)ethanone	<chem>O=C(c1ccc(N)cc1)C</chem>
2-methylaminobenzoic acid	<chem>OC(=O)c1c(NC)cccc1</chem>
2,5-dimethylphenol	<chem>Oc1c(ccc(c1)C)C</chem>
2,6-dimethylphenol	<chem>Oc1c(cccc1C)C</chem>
3,4-dimethylphenol	<chem>Oc1cc(c(cc1)C)C</chem>
(endo,endo)-7-oxabicyclo(2.2.1)heptane-2,3-dicarboxylic acid	<chem>O=C(O)[C@H]1[C@@H](C(=O)O)[C@@H]2O[C@H]1CC2</chem>
2-chloroprop-2-enyl diethylaminomethanedithioate	<chem>ClC(=C)CSC(=S)N(CC)CC</chem>
6-methylheptane-2,4-dione	<chem>C(C(=O)CC(=O)C)C(C)C</chem>
trans-1,4-dimethylcyclohexane	<chem>C[C@@H]1CC[C@H](CC1)C</chem>
1,1-dimethylcyclohexane	<chem>Cl(CCCCC1)(C)C</chem>
propylcyclopentane	<chem>C(CClCCCC1)C</chem>
(2s)-2-(2,4-dichlorophenoxy)propanoic acid	<chem>Clc1c(O[C@@H](C)C(=O)O)ccc(Cl)c1</chem>
2-(4-chloro-2-methylphenoxy)acetic acid	<chem>Clc1cc(c(OCC(=O)O)cc1)C</chem>
3-phenylpropanoic acid	<chem>OC(=O)CCc1ccccc1</chem>
methyl 4-methoxybenzoate	<chem>c1cc(ccc1OC)C(=O)OC</chem>
(3-methylphenyl) n-methylcarbamate	<chem>O(c1cc(ccc1)C)C(=O)NC</chem>
levodopa	<chem>Oc1cc(C[C@H](N)C(=O)O)ccc1O</chem>
2,3-dimethyl-1,2,3-butanetricarboxylic acid	<chem>OC(=O)[C@@](C(C)(C)C(=O)O)(CC(=O)O)C</chem>
nonanedioic acid	<chem>C(CCCCCC(=O)O)C(=O)O</chem>
octyl carbamate	<chem>O(CCCCCCCC)C(=O)N</chem>
chlordene	<chem>[C@@]12(Cl)C([C@]2([C@]([C@@H]2[C@H]1CC=C2)([C]Cl)Cl)Cl)Cl</chem>
5-amino-4-bromo-2-phenylpyridazin-3-one	<chem>BrC1C(=O)n(nc1N)c1ccccc1</chem>
4-amino-n-pyrazin-2-ylbenzenesulfonamide	<chem>S(=O)(=O)(Nc1ncnc1)c1ccc(N)cc1</chem>
(2s)-2-(4-chloro-2-methylphenoxy)propanoic acid	<chem>Clc1cc(c(O[C@@H](C)C(=O)O)cc1)C</chem>
5,6,7,8-tetrahydronaphthalen-2-ol	<chem>c1(O)cc2c(cc1)CCCC2</chem>
1-(4-chlorophenyl)sulfonyl-3-propylurea	<chem>Clc1ccc(S(=O)(=O)NC(=O)NCCC)cc1</chem>
2-methyl-5-propan-2-ylphenol	<chem>CC(C)c1cc(O)c(cc1)C</chem>
2-(2-hydroxyethyl-phenylamino)ethanol	<chem>C(N(CCO)c1ccccc1)CO</chem>
(1r,4s)-fenchone	<chem>O=C1[C@]2(C)[C@@H](Cl(C)C)CC2C</chem>
(-)-dihydrocarvone	<chem>O=C1[C@H](CC[C@@H](Cl)C(=C)C)C</chem>
2,7-dimethylquinoline	<chem>n1c2c(ccc(c2)C)ccc1C</chem>
4-amino-n-(4-methylpyrimidin-2-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1nc(ccn1)C)c1ccc(N)cc1</chem>
4-amino-n-(4,5-dimethyl-1,3-oxazol-2-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1oc(c(n1)C)c1ccc(N)cc1</chem>
s-ethyl (bis(2-methylpropyl)amino)methanethioate	<chem>S(C(=O)N(CC(C)C)CC(C)C)CC</chem>
benzene-1,4-diol; cyclohexa-2,5-diene-1,4-dione	<chem>c1cc(ccc1O).O.C1=CC(=O)C=CC1=O</chem>
1-[(2s)-3-butyn-2-yl]-3-(4-chlorophenyl)-1-methylurea	<chem>Clc1ccc(NC(=O)N([C@@H](C)C#C)C)cc1</chem>
4-amino-n-(2,6-dimethylpyrimidin-4-yl)benzenesulfonamide	<chem>S(=O)(=O)(Nc1nc(nc1)C)C)c1ccc(N)cc1</chem>
4-amino-n-(5-tert-butyl-1,3,4-thiadiazol-2-yl)benzenesulfonamide	<chem>s1c(nnc1NS(=O)(=O)c1ccc(N)cc1)C(C)(C)C</chem>
1-butyl-3-(4-methylphenyl)sulfonylurea	<chem>S(=O)(=O)(NC(=O)NCCCC)c1ccc(cc1)C</chem>
n,n-dibutylbutan-1-amine	<chem>N(CCCC)(CCCC)CCCC</chem>
1,3-di(phenyl)urea	<chem>N(C(=O)Nc1ccccc1)c1ccccc1</chem>
(6-tert-butyl-3-methyl-2,4-dinitrophenyl) acetate	<chem>O(c1c(C(C)(C)C)cc(N(=O)=O)c(c1N(=O)=O)C)C(=O)C</chem>

methyl dodecanoate	<chem>O(C=O)CCCCCCCCCCCCC</chem>
1-chloro-4-[2,2-dichloro-1-(2-chlorophenyl)ethenyl]benzene	<chem>Clc1c(C=C(Cl)Cl)c2ccc(Cl)cc2)cccc1</chem>
1,2-dihydroxyanthracene-9,10-dione	<chem>c12c(C(=O)c3c(Cl=O)ccc(c3O)O)cccc2</chem>
n-[(4-chlorophenyl)carbamoyl]-2,6-difluorobenzamide	<chem>Clc1ccc(NC(=O)NC(=O)c2c(F)cccc2F)cc1</chem>
1-chloro-2-[(1r)-2,2-dichloro-1-(4-chlorophenyl)ethyl]benzene	<chem>Clc1ccc(cc1)[C@@H](C(Cl)Cl)c1c(Cl)cccc1</chem>
2-[[3-(trifluoromethyl)phenyl]amino]benzoic acid	<chem>FC(F)(F)c1cc(Nc2c(cccc2)C(=O)O)ccc1</chem>
anthracen-9-ol	<chem>Oc1c2c(cc3c1cccc3)cccc2</chem>
2-(2-carboxyphenyl)benzoic acid	<chem>OC(=O)c1c(c2c(cccc2)C(=O)O)cccc1</chem>
4-(6-methyl-1,3-benzothiazol-2-yl)aniline	<chem>s1c2c(nc1c1ccc(N)cc1)ccc(c2)C</chem>
(1s)-1,2-diphenylethanol	<chem>O[C@@H](Cc1ccccc1)c1ccccc1</chem>
[(3-chloro-4-methyl-2-oxo-2h-chromen-7-yl)oxy] (diethoxy)sulfanylp phosphonium	<chem>Clc1c(c2c(oc1=O)cc(OP([S]))(OCC)OCC)cc2)C</chem>
n,n'-dimethyl-n-pyridin-2-yl-n- (thiophen-2-ylmethyl)ethane-1,2-diamine	<chem>s1c(CN(CCN(C)C)c2ncccc2)ccc1</chem>
1-(2-methylcyclohexyl)-3-phenylurea	<chem>O=C(N[C@@H]1[C@@H](CCCC1)C)Nc1ccccc1</chem>
2-[(2,3-dimethylphenyl)amino]benzoic acid	<chem>OC(=O)c1c(Nc2c(c(ccc2)C)C)cccc1</chem>
7-methoxy-8-(3-methylbut-2-enyl)chromen-2-one	<chem>C(C=C(C)C)c1c2c(ccc1OC)ccc(=O)o2</chem>
2-chloro-n-(2-ethyl-6-methylphenyl)-n- [(2r)-1-methoxy-2-propanyl]acetamide	<chem>ClCC(=O)N([C@@H](COC)C)c1c(CC)cccc1C</chem>
pentadecanoic acid	<chem>OC(=O)CCCCCCCCCCCCCCC</chem>
3-[4-(4-methoxyphenoxy)phenyl]-1,1-dimethylurea	<chem>O(c1ccc(NC(=O)N(C)C)cc1)c1ccc(OC)cc1</chem>
hexadecan-1-ol	<chem>OCCCCCCCCCCCCCCCCC</chem>
(r)-fenarimol	<chem>O[C@@](c1cncnc1)(c1ccc(cc1)Cl)c1c(ccc1)Cl</chem>
napropamide-m	<chem>O([C@@H](C(=O)N(CC)CC)C)c1c2c(ccc1)cccc2</chem>
3-(5,6-dihydrobenzo[b][1]benzazepin-11-yl) -n-methylpropan-1-amine	<chem>N1(c2c(CCc3c1cccc3)cccc2)CCCN</chem>
meso-hexestrol	<chem>Oc1ccc([C@H]([C@@H](CC)c2ccc(O)cc2)CC)cc1</chem>
alfatradiol	<chem>O[C@H]1[C@@]2([C@H]([C@H]3[C@H](CC2)c2c(CC3)cc(O)cc2)CC1)C</chem>
cinchonine	<chem>O[C@H]([C@@H]1N2C[C@@H]([C@H](Cl)CC2)C=C)c1c2c(nc1)cccc2</chem>
amitriptyline	<chem>N(CCC=Clc2c(CCc3c1cccc3)cccc2)(C)C</chem>
2-chloro-10-[3-(4-methylpiperazin-1-yl)propyl]phenothiazine	<chem>Clc1cc2N(CCCN3CCN(CC3)C)c3c(Sc2cc1)cccc3</chem>
quinine	<chem>O[C@@H]([C@H]1N2C[C@@H]([C@H](Cl)CC2)C=C)c1c2c(nc1)ccc(OC)c2</chem>
2-[5-methoxy-2-methyl-1-[(e)-3-phenylprop- 2-enoyl]indol-3-yl]acetic acid	<chem>O(c1cc2c(c(n(c2cc1)C(=O)/C=C/c1ccccc1)C)CC(=O)O)C</chem>
(1r)-trans-permethrin	<chem>ClC(=C[C@H]1C([C@@H]1C(=O)OC)c1cc(Oc2ccccc2)ccc1)(C)C)Cl</chem>
periciazine	<chem>S1c2c(N(CCCN3CCC(O)CC3)c3c1cccc3)cc(cc2)C#N</chem>
ethyl biscoumaracetate	<chem>olc(=O)c(C(c2c(O)c3c(oc2=O)cccc3)C(=O)OCC)c(O)c2c1cccc2</chem>

Cuadro 7.3 – Nombres y SMILES de las moléculas de la base de datos de 276 compuestos.

Capítulo 8

Aspectos prácticos del cálculo de los descriptores

8.1. Introducción

En su conjunto, las bases de datos empleadas contienen un total de $162 + 277 + 496 = 935$ especies químicas. Para realizar los cálculos de propiedades magnéticas es necesario el conocimiento de la geometría molecular de equilibrio de cada una de ellas. Para conocer la estructura de equilibrio de una especie química normalmente se realiza una optimización de geometría a partir de alguna razonable geometría de no equilibrio. El hecho de escribir estas geometrías es un problema en sí mismo. Este comprende, en primera instancia, averiguar los átomos y conectividades de cada molécula. Lo anterior debe hacerse a partir del dato identificador de las mismas provistos en las distintas bases de datos. Este proceso requiere búsqueda bibliográfica con excepción de casos triviales, es decir, cuando el identificador sea el nombre o la fórmula molecular de una sustancia ampliamente conocida.

Una vez conocidos los átomos y conectividades propios una molécula, se procede a dibujar la misma en algún programa de visualización de estructuras químicas. Hecho esto se exportan las coordenadas nucleares en algún formato conveniente.

Las coordenadas nucleares así obtenidas son ingresadas en la entrada de algún programa para cálculos cantidades que dependan de la estructura electrónica. Con este se ejecuta el proceso iterativo de optimización de geometría¹. El mismo no siempre es exitoso. Los

¹Utilizando como criterio la minimización de la suma entre la energía electrónica y la energía de repulsión núcleo-núcleo, expresada como función de las coordenadas nucleares.

motivos de falla que interesan aquí son tres: uno es que el punto estacionario, alcanzado al finalizar el proceso, puede ser un mínimo no global, lo cual es indeseable. Otro motivo de interés es que el punto estacionario alcanzado puede no ser un mínimo². Por último, el proceso podría progresar indefinidamente³.

Recién una vez optimizada la geometría, se procedería al cálculo de las propiedades magnéticas.

8.2. Descripción de la metodología empleada

Completar estos cálculos en tiempo razonable requiere la automatización de al menos algunos procesos. Estos procesos fueron clasificados en pasos que están descriptos con mayor detalle en las siguientes subsecciones.

8.2.1. Paso 1: Obtención de geometrías de partida

La identificación de las moléculas en las bases de datos utilizadas está dada por sus nombres o bien sus CAS Registry Number[®] CAS RN[®]⁴. Varias bases de datos permiten obtener información estructural de moléculas. Algunas de las primeras son accesibles vía Internet. Para evitar la gran demora que implicaría la búsqueda manual de estas estructuras, se desarrolló un programa informático que automatiza la tarea.

El programa fue desarrollado en el lenguaje Python3[57], y funcionaba de la siguiente manera. Leía desde un archivo de texto una lista de identificadores de especies químicas. Cada identificador era luego buscado en la base de datos Chemistry WebBook de NIST[58]. En caso de encontrarse la estructura, el programa devolvía la geometría molecular en coordenadas cartesianas. Como preferencia, cuando estaba accesible, el programa devolvía la estructura de la especie química en cuestión, obtenida esta mediante una optimización de geometría realizada con el método y base de interés (B3LYP/6-31G(d,p)). En algunos casos el programa encontraba más de una estructura para un dado identificador. En tales casos requería la inspección del usuario para la selección de la estructura apropiada. En caso de no encontrar la estructura, procedía a buscar en la bases de datos PubChem[59]. Todas las estructuras fueron encontradas de esta forma.

²En tal caso es un punto silla.

³O un máximo de iteraciones establecido para dicho proceso.

⁴Es un número de identificación único para cada molécula contenida la base de datos CAS REGISTRYSM.
Enlace al Sitio Oficial (funcional sólo para la versión electrónica).

8.2.2. Paso 2: Viabilidad de las geometrías de partida

La viabilidad de las geometrías de partida fue verificada. Esto resultó conveniente porque la inspección de las bases de datos reveló que algunas estructuras eran de baja calidad. Por ejemplo, estructuras planas para moléculas que no lo son.

Se escribió código en Python3 para verificar todas las estructuras. Con él se analizó, en primer lugar, la planaridad de las estructuras. En caso que los núcleos pudiesen ubicarse sobre un plano, el programa lanzaba un mensaje de alerta. Naturalmente se otorgó una pequeña tolerancia a la no planaridad.

El plano fue obtenido de una forma poco elegante. Se buscó a un par de núcleos cuya distancia mutua sea mayor o igual que a la correspondiente a cualquier otro par. Se encontró la recta que los atraviesa. Se escogió un tercer núcleo buscando maximizar la distancia a la recta. El plano elegido fue aquel que comprende a estos tres núcleos.

El programa también fue utilizado para analizar la conectividad de la molécula. Llámese fragmento al conjunto que contiene a un dado núcleo y todos aquellos núcleos accesibles al primero mediante el recorrido de enlaces covalentes. El programa lanzaba un mensaje de advertencia en caso de existir más de un fragmento en la estructura. La existencia de múltiples fragmentos no es necesariamente un error, pero se decidió evitar trabajar con esta clase de sustancias por resultar más difíciles de optimizar. Se descartó así al ítem 445 de la base de datos de 496 especies (CAS:106-34-3).

El algoritmo utilizado para el conteo de fragmentos se describe a continuación. Inicialmente, se estableció para cada par de núcleos la existencia o no de un enlace entre ellos. Para esto se consideró que: dos núcleos están enlazados sí cuando la distancia entre sus coordenadas es menor o igual que 1,2 veces la suma de sus radios covalentes. En dicho caso se los llamará vecinos. Después partiendo de un núcleo arbitrario, se lo incluye en un conjunto, y luego a sus vecinos. Los vecinos de los vecinos también son incluidos y así sucesivamente. Si al terminar el proceso el conjunto tiene menos elementos que el número de núcleos de la molécula, entonces la molécula tiene más de un fragmento.

8.2.3. Paso 3: Primera optimización de geometría

Las coordenadas cartesianas de los núcleos de las estructuras aceptadas habían sido guardadas en archivos con formato XYZ File. Se creó un pequeño *script* para generar los archivos de entrada para cálculos con Gaussian 09. Los mismos requerían al programa la optimización de geometría y el posterior cálculo de frecuencias vibracionales armónicas. Esto último para verificar, a partir de las mismas, que el punto estacionario encontrado

es un mínimo y no un punto silla. Se creó otro *script* para procesar estos archivos en el *cluster* del grupo. Los cálculos se realizaron al nivel de teoría B3LYP 6-31G(d,p).

Muchas estructuras convergieron a puntos silla. Para obtener los mínimos buscados se creó y aplicó un *script* cuyo funcionamiento es el siguiente. De cada archivo de salida lee la geometría, las frecuencias de vibración armónica y las contribuciones, en coordenadas cartesianas, de los modos normales de vibración. Para cada modo normal cuya frecuencia de vibración no sea positiva, genera un desplazamiento en la geometría de su correspondiente contribución multiplicada por un factor de escala. Luego de unos pocos ensayos se decidió otorgar el valor 0.25 a este factor.

Una vez perturbadas las geometrías se repitió el proceso de optimización. Fue necesario aplicar dos veces este procedimiento a algunas moléculas. De este modo todas las geometrías obtenidas resultaron ser mínimos de las superficies de energía potencial correspondientes.

8.2.4. Paso 4: Estudio conformacional

Fue decidido en esta instancia realizar un estudio conformacional de cada una de las especies químicas en cuestión. Para que esto pueda ser realizado se desarrolló una serie de *scripts* que, sobre cada molécula, realizan lo siguiente. En primera instancia generan un muestreo de rotámeros creados a partir de la estructura de equilibrio obtenida previamente. Luego verifican la validez de las estructuras generadas y descartan aquellas que no sean de utilidad. A continuación generan las entradas para optimizar la geometría de los rotámeros con el método semiempírico PM7[60] implementado en el programa Mopac 2012[61]. Para cada especie química escogen una estructura entre las de menor energía, y generan los archivos de entrada para una optimización al nivel B3LYP/6-31G(d,p). Estas estructuras son comparadas con las obtenidas en el Paso 3, y reemplazan a las primeras sólo en caso de ser de menor energía.

Generación de rotámeros

Para cada molécula con N enlaces rotables, se generó $\min[3^N, 729]$ rotámeros a partir de su geometría de equilibrio previamente encontrada.

En moléculas con hasta seis enlaces rotables, por cada uno de ellos se utilizó tres ángulos diedros diferenciados en $2/3\pi$ rad. En cada trío de ángulos, uno corresponde al de la geometría de equilibrio previamente hallada. Se generó las 3^N estructuras posibles a partir de combinaciones de estos ángulos. Estas fueron guardadas en archivos con formato

XYZ File.

En moléculas con más de seis enlaces rotables, sólo se generó $3^6 = 729$ estructuras. Estas fueron escogidas en forma aleatoria entre todas aquellas obtenibles con el procedimiento anterior. Esta limitación se aplicó para restringir la cantidad de rotámeros a un número manejable.

Dada la dificultad para la detección de enlaces rotables, se utilizó a la biblioteca OpenBabel[62] y a Pybel[63]. La primera tiene incorporada esta función, la segunda simplifica el trabajo con la primera.

El funcionamiento general del programa es el siguiente: Se itera sobre cada molécula. Para cada molécula se detectan los enlaces rotables. En caso de no existir, continua con la próxima molécula. En caso de existir, guarda los pares de núcleos involucrados en cada enlace rotatable. Para cada enlace rotatable busca (dos) núcleos vecinos que permitan definir un ángulo diedro. Obtiene cada ángulo y calcula los ángulos correspondientes a dos rotaciones en $2/3\pi$ rad. Genera una lista con todas las combinaciones posibles de los distintos ángulos diedros considerados. Para hacer esto en forma eficiente, se pensó la siguiente solución al problema: asociar un conjunto a cada enlace rotatable, donde cada conjunto contiene los tres ángulos diedros correspondientes. Luego interpretar a las combinaciones buscadas como elementos del conjunto obtenido mediante el producto cartesiano entre los conjuntos con tres elementos. Esto permite utilizar bibliotecas preexistentes bien optimizadas.

Si la lista tiene más que 729 elementos, entonces deben tomarse sólo 729 en forma aleatoria⁵. Para esto se implementó el algoritmo de Fisher y Yates [64] modificado por Durstenfeld[65], que ordena en forma no sesgada⁶ una lista finita.

Verificación de conectividades

Los archivos de las geometrías generadas en el paso anterior preservan el orden en que fueron listados los núcleos en el archivo a partir del cual se crearon. Esto fue realizado intencionalmente, para verificar que la conectividad de cada núcleo no haya sido modificada luego de aplicar las rotaciones.

El resultado de este proceso fue la eliminación de gran cantidad de estructuras. En la base de datos de 162 compuestos, se eliminaron 230 estructuras de 1548 totales. En la base

⁵El programa fue diseñado para que devuelva las estructuras en forma no completamente aleatoria, sino que siempre devuelva las mismas estructuras (fijando la semilla del generador de números pseudoaleatorios). Esto fue hecho así para aumentar la reproducibilidad de los resultados.

⁶Cada permutación es igualmente probable.

de datos de 277 moléculas, 9175 de 28905. En aquella de 496 especies, 19836 de 52743. Se eliminaron entonces 29241 de 83196 estructuras ($\approx 35\%$)

Obtención de confórmeros al nivel PM7

Las geometrías de los rotámeros obtenidos no corresponden, en principio, a geometrías de equilibrio. Dada la gran cantidad de rotámeros, se utilizó un método ligero para optimizar sus geometrías. Las energías de equilibrio obtenidas en dicho proceso se utilizaron para establecer el criterio de un primer filtro de estructuras poco relevantes. El método escogido fue PM7[60], que es de tipo semiempírico, y se lo aplicó utilizando el programa Mopac 2012[61].

Las geometrías de los confórmeros de la especie 119 de la base de datos de log P fueron reoptimizadas. Para ello se utilizó la funcional SVWN con la base 3-21G*. Las estructuras presentaron diferencias notables. Varias estructuras originalmente distintas con PM7 convergían a la misma estructura con SVWN/3-21G*. Surgió entonces el interés por utilizar un método alternativo a PM7 para la optimización de geometrías. Se probó el método DFTB The Density Functional based Tight Binding [66, 67, 68]. Los resultados fueron muy similares a los de PM7, con un costo computacional unas 20 veces mayor. Por tal motivo se descartó la idea.

Fiabilidad de las energías

En el paso anterior se obtuvieron geometrías de equilibrio y sus correspondientes energías. Es posible ordenar los confórmeros de acuerdo con las energías obtenidas, pero el ordenamiento puede no ser el correcto porque las energías fueron aproximadas. Se pretende ver cuán fiables son estos ordenamientos respecto de un método mejor.

Utilizando la molécula 119 mencionada se realizó una comparación entre los resultados obtenidos con PM7 y aquellos de SVWN/3-21G*. Para ambos conjuntos de resultados, se expresaron las energías relativas a sendos confórmeros de mínima energía. Para cada rotámero de origen se compararon las energías relativas de sendos métodos por inspección visual. Si bien hubo una correspondencia razonable, no fue exacta.

Se diseñó la siguiente comparación con fines cuantitativos. Para cada confórmero se obtuvo la diferencia entre ambas energías relativas. A este conjunto de datos se le sacó la media y desviación estándar. Estas fueron 0.56 y 0.43 kcal/mol. Con gran confianza se puede esperar que no sean óptimos verdaderos, aquellos confórmeros con energía mayor en $(0, 56 + 2 \times 0, 43)$ kcal/mol que aquel de menor energía a nivel PM7.

Lo anterior es válido sólo para la molécula en cuestión, y suponiendo que el método SVWN/3-21G* es adecuado. Estos números no debieran ser aplicables en forma directa a otras moléculas, en parte por las diferencias en naturaleza de los núcleos y tipos de enlace en cuestión. Puede ser más importante aún la diferencia en número de enlaces entre las especies.

Comentario 8.2.1. Los errores absolutos en la energía molecular, de esencialmente todo método computacional aplicable, son por en mucho mayores que las energías de un enlace químico. El éxito en las predicciones se debe esencialmente a una compensación de errores sistemáticos. Las compensaciones son mucho mayores en la zona cercana a los núcleos, dado que estas están menos afectas por los enlaces químicos. Los enlaces químicos son entonces la zona de mayor error y la que debe analizarse acá.

Para otras moléculas, como aproximación cruda se utilizó una tolerancia por enlace igual a $(0.56 + 2 \times 0.43)$ kcal/mol dividido entre el número de enlaces de la molécula 119 (20). Es decir, aquellas estructuras cuya energía sea superior a la energía mínima para la especie en $(0.56 + 2 \times 0.43)$ kcal/mol / 20 sería descartada. Se creó un pequeño *script* en Python para realizar esta tarea.

Varios miles de estructuras fueron aceptables bajo este criterio. Para realizar las optimizaciones en un tiempo razonable se requería un poder de cómputo no disponible. Por tal motivo se procedió de la siguiente manera: Primero escoger y optimizar una estructura entre las de menor energía⁷ al nivel B3LYP/6-31G(d,p)⁸. Luego compararla con la estructura original optimizada mismo nivel de teoría. Retener la estructura original salvo que la nueva estructura sea de menor energía. Para esto también se realizó un pequeño *script* en Python3.

8.2.5. Paso 5: Cálculo de las propiedades magnéticas

Una vez obtenidas las estructuras finales, las propiedades magnéticas fueron calculadas con G09. Se utilizó el nivel de teoría B3LYP/6-311+G(2d,p) recomendado por Cheesman[30], y se utilizó el método GIAO[69, 70, 71, 72]. En cada caso, la línea de entrada utilizada fue

```
#p B3LYP/6-311+G(2d,p) NMR(GIAO, PrintEigenvectors,  
Susceptibility,SpinSpin) SCRF=(Solvent=Water)
```

⁷Se está considerando el caso degenerado.

⁸Las optimizaciones de geometrías se realizaron en la misma forma que en el Paso 3.

El significado de las opciones del bloque NMR es: GIAO, invoca al método GIAO. SpinSpin para el cálculo de las constantes espín-espín. Susceptibility para el cálculo de la susceptibilidad magnética⁹.

Los cálculos tardaron entre minutos y días dependiendo de la molécula en cuestión, y las computadoras utilizadas. Algunos de ellos fallaron según se lo describió en el capítulo de bases de datos.

8.2.6. Cálculo de los descriptores

Los archivos con la información de cálculos de propiedades magnéticas fueron utilizados para la creación de descriptores.

Debido a las formas funcionales de los descriptores elegidos, descritas en la sección 5.2.1, se requiere valores únicos de cada núcleo para reemplazar a los parámetros A_i . Sin embargo, a veces la información magnética contiene propiedades referidas a un par de núcleos. Por ejemplo, en el caso de los tensores de acoplamiento, cada elemento no diagonal de sus representaciones matriciales contienen información de dos núcleos.

Esto no representa un problema porque, en la mayoría de los descriptores usados, la propiedades nucleares A_i aparecen en productos $A_i \cdot A_j$. En tales casos dichos productos pueden ser reemplazados, por ejemplo, con el elemento de matriz ij correspondiente. Ese fue el procedimiento usado en este trabajo. Nótese que los elementos diagonales son nulos en el caso mencionado.

Los tensores de apantallamiento magnético son específicos de cada núcleo. Sin embargo, sus representaciones matriciales contienen nueve elementos, lo cual impide el reemplazo directo del correspondiente A_i . Para obtener un único valor por núcleo, se tomó el cuadrado de la media aritmética de los tres autovalores de la representación matricial su tensor de apantallamiento. Dicho promedio es conocido como *apantallamiento magnético isotrópico*

Para este último paso se desarrolló código en Python3. El mismo está disponible en los apéndices de esta tesis. También un ejemplo de cómo obtener los A_i a partir de un archivo de salida de Gaussian 09. El resultado de la aplicación de este código son dos archivos sobre un conjunto de ficheros son dos nuevos archivos de texto. Uno contiene la

⁹La susceptibilidad magnética (χ) se define con

$$\mathbf{M} = \chi \mathbf{B}$$

donde \mathbf{M} representa la magnetización, es decir, el momento dipolar magnético $\boldsymbol{\mu}$ (definido por $\boldsymbol{\tau} = \boldsymbol{\mu} \times \mathbf{B}$) por unidad de volumen.

lista de los nombres de los descriptores creados. El otro contiene una matriz donde cada fila corresponde a una molécula y cada columna a un descriptor de la misma.

La nomenclatura utilizada en los archivos es determinada por las siguientes reglas. La cadena previa al primer guión bajo representa a los tipos de interacciones magnéticas: f c al contacto de Fermi, sd a espín dipolar, pso y dso a paramagnéticas espín-órbita y diamagnéticas espín-órbita respectivamente, y tnss a espín-espín nuclear total. Entre ambos guiones bajos, j referencia al tensor de acoplamiento indirecto espín-espín, y k a la versión reducida. El número entero final corresponde al parámetro arbitrario, excepto en los siguientes casos: en el caso de los descriptores GETAWAY, 0, 1 y dos significa HATS(A), HT(A) and RT(A) respectivamente. En el caso de los descriptores WHIM 0 y 1 significan T y A respectivamente, mientras que 2, 3 and 4 se refieren a los θ_i s en orden ascendente.

Cálculo de descriptores con eDragon

Con fines comparativos se calcularon los descriptores utilizando al programa eDragon[54]. Para ello se convirtieron los archivos de salida de G09 al formato de moléculas múltiples SDF, y se utilizó la plataforma digital de eDragon para su cálculo.

Parte V

RESULTADOS

Capítulo 9

Viabilidad de los descriptores

Una vez obtenidos exitosamente los descriptores, se desea corroborar que ellos puedan ser relacionados con otras propiedades cuantificables. La calidad de las correlaciones fueron comparadas con aquellas obtenibles para los descriptores calculados con eDragon.

Utilizando estos descriptores, los nuevos y los de eDragon, se crearon modelos para las tres bases de datos descriptas anteriormente. Para cada base de datos, se crearon modelos utilizando el ERM y distinto número de descriptores; entre 1 y 10. Para cada número de descriptores y base de datos, se crearon tres conjuntos de modelos; aquellos que incluyen sólo descriptores nuevos, otros con descriptores de eDragon exclusivamente, y otro en el que ambos conjuntos de descriptores fueron puestos a disponibilidad.

La calidad del modelo suele ser representado por dos parámetros diferentes, la desviación estándar S de los residuos, y el coeficiente de correlación de Pearson R entre la propiedad predicha y su valor aceptado. Dado que existe una gran correlación entre ambos, se decidió utilizar sólo uno de ellos en este reporte comparativo.

Para cada base de datos y conjunto de origen de los descriptores, se obtuvieron los parámetros estadísticos más significativos para cada modelo. Entre los $3 \cdot d$ modelos creados para cada caso de d descriptores, se seleccionó aquel con menor valor de S_{TS} para ser incluido en las tablas presentadas más adelante. Los resultados para la base de datos de 160 moléculas están tabulados en la Tabla 9.1. Se tabularon los parámetros más significativos del mejor modelo encontrado, la desviación estándar en el conjunto de ajuste, la correspondiente de LOO (S_{loo}) y del conjunto de pruebas S_{TS} .

Naturalmente, dado que no se ha explorado todo el conjunto de combinaciones de descriptores (excepto para modelos de un descriptor), los resultados sólo tienen carácter indicativo. Incluso si hubiese sido posible realizar la búsqueda exhaustiva, la exactitud de los resultados estaría limitada al hecho de sólo utilizar tres conjuntos de moléculas y

	Nuevos			eDragon			Mezcla			
d	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	nuevos
1	0,82	0,83	0,63	0,77	0,79	0,72	0,78	0,80	0,70	0
2	0,68	0,70	0,61	0,53	0,55	0,53	0,49	0,51	0,47	1
3	0,56	0,60	0,49	0,45	0,46	0,55	0,45	0,47	0,47	2
4	0,45	0,49	0,52	0,38	0,40	0,46	0,38	0,39	0,37	2
5	0,41	0,44	0,47	0,36	0,37	0,45	0,34	0,36	0,38	3
6	0,37	0,40	0,42	0,32	0,34	0,42	0,32	0,35	0,37	3
7	0,35	0,39	0,35	0,31	0,33	0,42	0,29	0,33	0,36	3
8	0,32	0,35	0,39	0,28	0,31	0,40	0,27	0,29	0,34	5
9	0,29	0,34	0,37	0,25	0,29	0,42	0,25	0,26	0,37	2
10	0,26	0,29	0,38	0,26	0,29	0,39	0,26	0,29	0,41	6
<i>Suma</i>	3,96	4,22	3,88	3,39	3,55	3,95	3,32	3,50	3,47	

Cuadro 9.1 – Desviación estándar en el conjunto de ajuste *S*, de LOO (*S_{loo}*) y del conjunto de pruebas *S_{TS}* para los descriptores nuevos, de eDragon y la unión de ambos. *d* representa el número de descriptores en el modelo y “nuevos” el número de descriptores nuevos en el modelo.

tres propiedades moleculares diferentes.

Para determinar la existencia de correlación es útil comparar con algún resultado que dé idea de qué puede esperarse ante la ausencia de correlación. Según lo dicho en el Comentario 5.4.1, sería apropiado con fines comparativos usar la media como predictor. En este conjunto de 160 moléculas, la media es 1,60. La desviación estándar de los residuos así obtenidos es 1,40.

Comentario 9.0.1. Nótese que esta comparación tiene mucha semejanza con el coeficiente de determinación,

$$R^2 := 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

que tratándose de una regresión lineal

$$R^2 = r^2 = \frac{\sum_i (\bar{y} - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

Con esta comparación, los resultados hacen evidente que existe una correlación entre los descriptores desarrollados en este trabajo de tesis y la propiedad (Log P en este caso) y

que puede ser explotada considerándola lineal.

En cuanto al uso con fines predictivos, no resulta demasiado clara la conveniencia de alguno de los conjuntos de descriptores por sobre el otro. Una forma de comparación es a través de la suma de los valores de S , S_{loo} y S_{TS} para los distintos casos (Tabla 9.1). Comparar para un único d carece de sentido aquí dado que los resultados varían en gran medida con d . Si bien los descriptores de eDragon arrojaron mejores resultados en la validación interna (3,55 frente a 4,22), en el conjunto de prueba la tendencia fue revertida (3,95 frente a 3,88). El uso conjunto de ambos grupos de descriptores resultó comúnmente ventajoso.

Se debe tener presente que estos resultados fueron obtenidos utilizando sólo una pequeña fracción del total de combinaciones posibles de descriptores. No puede obviarse el hecho de que existe una desviación estándar significativa de los errores medios, para un mismo conjunto de moléculas y descriptores disponibles, entre las predicciones obtenidas con los distintos modelos creados mediante el algoritmo de selección empleado.

Los resultados correspondientes a la base de datos de 276 especies se encuentran en la tabla 9.2. En este caso, la media de la temperatura de ebullición es 401 K. Si se la usa como predictor se obtiene una desviación estándar de los residuos de 84 K. Los resultados confirman la correlación entre los descriptores propuestos y las temperaturas de ebullición. En este caso los resultados de eDragon son notoriamente superiores. El motivo se debe al anormalmente buen desempeño del descriptor X_{lsol} [73] para esta propiedad. Es destacable que se haya obtenido un mayor S_{TS} con el conjunto completo que con sólo los descriptores de eDragon.

Los resultados correspondientes a la base de datos de 453 especies se encuentran en la tabla 9.3. Como propiedad a predecir se utilizó el logaritmo decimal del valor numérico de la solubilidad acuosa expresada en gramos por litro. En este caso, usando la media (-2,85) como predictor se obtiene una desviación estándar de los residuos de 2,1. Nuevamente los resultados confirman la correlación entre los nuevos descriptores y una propiedad de interés, solubilidad acuosa en este caso.

Resulta de interés explorar la importancia de adicionar los descriptores propuestos a los de eDragon. En las tablas anteriores se indicó el número de descriptores nuevos incluidos en cada uno de los diez modelos empleados. Más información puede obtenerse mediante el análisis de todos los diferentes modelos creados. En cada modelo de d descriptores, se analizó la importancia relativa de cada uno de ellos y se los ordenó de mayor a menor importancia. Esta fue cuantificada en términos de los coeficientes de regresión estandarizados.

	Nuevos			eDragon			Mezcla			
d	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	nuevos
1	52,4	53,0	52,9	43,5	43,9	37,5	41,3	41,7	42,3	0
2	29,6	30,1	33,9	26,5	27,1	19,9	24,1	24,6	26,0	1
3	24,1	24,7	31,6	22,1	22,6	16,7	22,1	22,9	22,0	2
4	22,9	23,7	28,6	19,7	20,3	14,3	16,6	17,2	18,7	1
5	20,2	21,0	27,7	18,5	19,3	13,9	15,3	16,0	17,9	1
6	18,7	19,7	28,3	16,9	17,7	12,6	13,8	14,5	16,6	2
7	17,8	18,9	26,9	15,0	15,8	11,8	12,9	13,7	14,3	3
8	17,4	18,6	25,1	15,0	15,9	12,4	12,7	13,6	13,8	4
9	14,4	15,4	27,7	13,8	14,7	11,9	11,5	12,2	13,6	3
10	13,7	14,8	27,1	12,6	13,4	11,4	10,8	11,9	13,9	5
<i>Suma</i>	231,3	239,9	309,8	203,6	210,8	162,4	181,1	188,3	199,0	

Cuadro 9.2 – Desviación estándar en el conjunto de ajuste *S*, de LOO (*S_{loo}*) y del conjunto de pruebas *S_{TS}* para los descriptores nuevos, de eDragon y la unión de ambos. *d* representa el número de descriptores en el modelo y “nuevos” el número de descriptores nuevos en el modelo.

Comentario 9.0.2. Sea un modelo lineal cuya ecuación para la predicción de la propiedad de la molécula *i* (*y_i*) esté determinada por $\hat{y}_i = \alpha^T \cdot \mathbf{x} + \epsilon_i$, donde \mathbf{x} representa el vector de los valores de los descriptores para la especie *i*, $\alpha^T = [\alpha_1, \alpha_2, \dots, \alpha_d]$ la matriz fila de los coeficientes obtenidos para el modelo, y ϵ el error. Sus coeficientes estandarizados

$$\tilde{\alpha}_j := \alpha_j \frac{S_{x_j}}{S_y}$$

pueden interpretarse de la siguiente forma: Un cambio en *x_j* de magnitud igual a una desviación estándar, genera una variación en *y* de $\tilde{\alpha}_j$ desviaciones estándares.

En caso de existir un término independiente en el modelo, no se le asocia un coeficiente estandarizado. Esto se debe a que su coeficiente es 1 y por lo tanto su desviación estándar es nula.

Luego se contabilizó cuantas veces un descriptor nuevo fue el más importante, cuantas veces el segundo en importancia, etc. Esto se realizó para los tres conjuntos. Los resultados se encuentran en la Tabla 9.4.

Los datos tabulados permiten deducir existe una tendencia global a elegir a los descriptores de eDragon por sobre los propuestos. Los porcentajes de elección de descriptores nuevos fueron 48, 43 y 41 % según se muestra en la tabla. Estos porcentajes, si bien menores

d	Nuevos			eDragon			Mezcla			nuevos
	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	<i>S</i>	<i>S_{loo}</i>	<i>S_{TS}</i>	
1	1,45	1,46	1,37	1,45	1,46	1,48	1,44	1,45	1,49	1
2	1,27	1,28	1,20	1,14	1,16	0,97	1,11	1,12	1,15	0
3	1,08	1,10	0,98	0,99	1,01	0,77	0,90	0,91	0,81	0
4	0,98	1,00	0,93	0,84	0,86	0,76	0,79	0,81	0,76	0
5	0,95	0,99	0,89	0,79	0,81	0,69	0,74	0,75	0,69	1
6	0,88	0,90	0,80	0,72	0,73	0,68	0,70	0,72	0,67	2
7	0,82	0,85	0,78	0,68	0,70	0,65	0,67	0,69	0,63	3
8	0,80	0,82	0,77	0,69	0,72	0,64	0,66	0,68	0,62	3
9	0,76	0,79	0,75	0,64	0,66	0,60	0,63	0,66	0,61	3
10	0,75	0,78	0,74	0,63	0,67	0,62	0,59	0,61	0,63	4
Suma	8,22	8,39	7,72	7,30	7,44	6,64	7,01	7,14	6,81	

Cuadro 9.3 – Desviación estándar en el conjunto de ajuste *S*, de LOO (*S_{loo}*) y del conjunto de pruebas *S_{TS}* para los descriptores nuevos, de eDragon y la unión de ambos. *d* representa el número de descriptores en el modelo y “nuevos” el número de descriptores nuevos en el modelo.

que los correspondientes a eDragon, son considerables.

9.1. Modelo para predicción de Log P

Habiendo establecido correlación entre los nuevos descriptores y las propiedades, un modelo para la predicción de $\log P$ fue creado con fines demostrativos. Para ello se utilizó la base de datos de 160 moléculas anteriormente descrita. Se usó un modelo de siete descriptores ($d = 7$), elegidos entre un total de 1094 descriptores por molécula contruidos ($D = 1094$). El método de obtención de descriptores usado es el método de reemplazo ampliado del Dr. Mercader. El modelo encontrado fue

$$\begin{aligned} \log P_{\text{estimado}} = & -132,993(\pm 6) + 0,4667(\pm 0,1)\text{dso_j-rdf9} \\ & - 0,1559(\pm 0,02)\text{dso_j-rdf13} - 1,7210(\pm 0,2)\text{dso_k-ats2} \\ & - 4,9375(\pm 0,4)\text{dso_k-rdf23} + 2,4135(\pm 0,1)\text{pso_j-morse3d29} \\ & + 1,7893(\pm 0,1)\text{sd_j-morse3d6} + 3,3881(\pm 0,2)\text{sd_j-morse3d15} \end{aligned} \quad (9.1)$$

Posición	1º	2º	3º	4º	5º	6º	7º	8º	9º	10º	Total
log P											
eDragon	37	88	56	76	64	65	61	47	33	8	535
Nuevos	97	45	74	52	59	47	36	32	23	22	487
%Nuevos	72 %	34 %	57 %	41 %	48 %	42 %	37 %	41 %	41 %	73 %	48 %
Puntos de ebullición											
eDragon	101	80	21	82	65	42	33	47	24	15	510
Nuevos	18	38	96	33	46	59	52	20	21	9	392
%Nuevos	15 %	32 %	82 %	29 %	41 %	58 %	61 %	30 %	47 %	38 %	43 %
log (solubilidad acuosa)											
eDragon	109	85	73	70	57	52	38	29	20	14	547
Nuevos	11	34	45	44	53	51	55	45	33	13	384
%Nuevos	9 %	29 %	38 %	39 %	48 %	50 %	59 %	61 %	62 %	48 %	41 %

Cuadro 9.4 – Cantidad de veces en las que algún descriptor del conjunto (eDragon o Nuevos) fue elegida en los distintos órdenes de importancia. %Nuevos indica el porcentaje de veces que fue elegido un descriptor del conjunto Nuevos. Total es la cantidad total de veces que los descriptores fueron elegidos.

$$\begin{aligned}
 N &= 107 & FIT &= 13,606 & p < 10^{-5} \\
 R &= 0,9775 & S &= 0,3220 \\
 R_{loo} &= 0,9725 & S_{loo} &= 0,3551 \\
 R_{l-30\%-o} &= 0,9553 & S_{l-30\%-o} &= 0,4718 \\
 R_{TS} &= 0,9421 & S_{TS} &= 0,3745
 \end{aligned}$$

donde los errores estándares de los coeficientes de regresión fueron escritos entre paréntesis. N representa al número de moléculas incluidas en el conjunto de ajuste¹. p representa la significancia del modelo y FIT a la función de Kubinyi². El superíndice TS representa asociación con el conjunto de prueba. El significado de los nombres de los descriptores, usados en el modelo (Eq. 9.1), se encuentra en la tabla 9.6. Los elementos de la matriz de correlación de los descriptores pueden encontrarse en la tabla 9.5. De ellos se deduce que no hay un grado relevante de intercorrelación.

¹Aproximadamente un 30 %.

²Es una medida de la bondad del ajuste, definida como

$$FIT = \frac{R^2(N - d - 1)}{(N + d^2)(1 - R^2)}$$

	dso_j-rdf-13	dso_k-ats-2	dso_k-rdf-23	pso_j-morse3d-29	sd_j-morse3d-6	sd_j-morse3d-15
dso_j-rdf-9	0.2409	0.1168	0.5867	0.2530	0.1562	0.2261
dso_j-rdf-13	1	0.0180	0.0255	0.2626	0.2166	0.0664
dso_k-ats-2		1	0.2490	0.4418	0.3853	0.2806
dso_k-rdf-23			1	0.2922	0.1510	0.6607
pso_j-morse3d-29				1	0.4949	0.5012
sd_j-morse3d-6					1	0.1339

Cuadro 9.5 – Matriz de correlación entre los descriptores incluidos en el modelo descrito por la ec. 9.1. Los valores en negritas representan la mayor correlación. Para conservar espacio, se eliminó la información redundante.

Descriptor Name	Definition
dso_j-rdf-9	Contribución diamagnética espín-órbita a J , usando la familia RDF con $R = 9$
dso_j-rdf-13	Contribución diamagnética espín-órbita a J , usando la familia RDF con $R = 13$
dso_k-ats-2	Contribución diamagnética espín-órbita a K , usando la familia ATS con $k = 2$
dso_k-rdf-23	Contribución diamagnética espín-órbita a K , usando la familia RDF con $R = 23$
pso_j-morse3d-29	Contribución paramagnética espín-órbita a J , usando 3D-Morse con $s = 29$
sd_j-morse3d-6	Contribución espín dipolar a K , usando 3D-Morse con $s = 6$
sd_j-morse3d-15	Contribución espín dipolar a K , usando 3D-Morse con $s = 15$

Cuadro 9.6 – Descriptores del modelo descrito por la ec. 9.1.

Name	Exp.	Calc.	res.	Name	Exp.	Calc.	res.
Naphthalen-2-amine	2.28	2.60	-0.32	Phenylmethanamine	1.09	1.34	-0.25
Acetic acid ^{TS}	-0.17	-0.10	-0.07	Benzamide	0.64	0.64	0.00
1H-Imidazole ^{TS}	-0.08	0.27	-0.35	N-Methylbenzamide	0.86	1.45	-0.59
Propanoic acid ^{TS}	0.33	0.45	-0.12	N-Methylaniline ^{TS}	1.66	2.45	-0.79
Aniline ^{TS}	0.90	1.07	-0.17	Phenylcyanamide ^{TS}	1.87	1.52	0.35
Phenylmethanol	1.10	1.03	0.07	N-Phenylacetamide	1.16	1.12	0.04
2-Hydroxybenzamide	1.28	0.37	0.91	N-Phenylhydroxylamine ^{TS}	0.79	0.76	0.03
4-Hydroxybenzoic acid	1.56	1.34	0.22	N-Phenylaniline	3.50	3.67	-0.17
2,4-Dihydroxybenzoic acid	1.64	1.47	0.17	Phenol	1.46	1.50	-0.04
Benzoic acid	1.87	1.67	0.20	ethanol	-0.31	-0.16	-0.15
2-(3-methylphenyl)acetic acid	1.95	2.43	-0.48	Methanamine	-0.57	-0.71	0.14
3-Methoxybenzoic acid	2.02	2.15	-0.13	Methanol ^{TS}	-0.77	-0.97	0.20
N,N-dihydroxy-4-methylaniline	2.37	2.54	-0.17	Octan-1-ol	3.00	3.22	-0.22
4-Phenylbutanoic acid	2.42	2.51	-0.09	Ethanamine	-0.13	0.01	-0.14
2-Methylbenzoic acid ^{TS}	2.46	2.12	0.34	N-Methylmethanamine ^{TS}	-0.38	0.97	-1.35
Naphthalen-1-ol	2.85	3.15	-0.30	Cyclohexanol	1.23	1.37	-0.14
Di(phenyl)methanone	3.18	3.06	0.12	N-ethyl-1-phenylcyclohexan-1-amine	1.49	2.22	-0.73
Naphtalene	3.30	3.48	-0.18	2H-triazole	-0.29	-0.51	0.22
4-Butylbenzoic acid	3.97	3.75	0.22	1H-1,2,4-Triazole ^{TS}	-0.58	-0.41	-0.17
1,2,3,4,5-Pentamethylbenzene	4.56	4.68	-0.12	4-Nitroaniline	1.39	0.92	0.47
2-phenylethylbenzene	4.79	4.63	0.16	3-Nitroaniline	1.37	1.15	0.22
2-Methylphenanthrene	4.86	4.68	0.18	2-Nitroaniline ^{TS}	1.85	1.56	0.29
Pyrene	4.88	4.64	0.24	Benzene-1,4-diol	0.59	0.95	-0.36
2-Ethylanthracene	5.85	5.59	0.26	Benzene-1,3-diol ^{TS}	0.80	0.85	-0.05
Benzene ^{TS}	2.13	2.00	0.13	benzene-1,2-diol	0.88	0.93	-0.05
Methylbenzene ^{TS}	2.73	2.57	0.16	4-Nitrophenol	1.91	1.72	0.19
Ethylbenzene	3.15	3.01	0.14	3-Nitrophenol	2.00	1.71	0.29
Ethenylbenzene	2.95	2.93	0.02	2-(Dihydroxyamino)phenol ^{TS}	1.79	2.03	-0.24
Ethynylbenzene ^{TS}	2.53	3.10	-0.57	4-Nitrobenzoic acid	1.89	1.80	0.09
Propylbenzene	3.72	3.46	0.26	3-Nitrobenzoic acid ^{TS}	1.83	1.25	0.58
Prop-2-enylbenzene	3.23	3.72	-0.49	2-Nitrobenzoic acid ^{TS}	1.46	1.48	-0.02
Isopropylbenzene	3.66	3.60	0.06	4-Aminobenzoic acid ^{TS}	0.83	0.70	0.13
Butylbenzene ^{TS}	4.38	4.31	0.07	4-Aminophenol	0.04	0.45	-0.41
Tert-Butylbenzene	4.11	3.44	0.67	4-Cyanobenzoic acid ^{TS}	1.56	1.49	0.07
Cyclopropylbenzene	3.27	3.65	-0.38	4-Hydroxybenzonitrile ^{TS}	1.60	1.51	0.09
Phenylbenzene	4.01	4.29	-0.28	4-Methoxybenzoic acid	1.96	2.15	-0.19
phenylmethylbenzene	4.14	4.21	-0.07	2-Methylpyridine	1.11	1.18	-0.07
Ethane	1.81	1.20	0.61	2-Pyridinecarbonitrile ^{TS}	0.40	0.52	-0.12
Ethene	1.13	0.75	0.38	Ethyl pyridine-2-carboxylate	0.87	1.30	-0.43
Acetylene	0.37	0.39	-0.02	Methyl pyridine-2-carboxylate	0.43	0.75	-0.32
Propane ^{TS}	2.36	2.00	0.36	N,N-Dimethylpyridin-2-amine	1.65	1.56	0.09
Prop-1-ene ^{TS}	1.77	1.76	0.01	2-Ethoxypyridine ^{TS}	1.81	2.08	-0.27
Prop-1-yne	0.94	1.02	-0.08	2-Methoxypyridine	1.34	1.22	0.12
Butane	2.89	2.40	0.49	1-(2-Pyridyl)ethanone	0.83	0.46	0.37
2-Methylpropane	2.76	2.70	0.06	Pyridin-2-ylmethyl acetate	0.39	0.61	-0.22
Buta-1,3-diene	1.99	2.30	-0.31	2-Ethylpyridine ^{TS}	1.60	1.85	-0.25
Pentane	3.39	3.00	0.39	2-Phenylpyridine	2.65	2.51	0.14
2,2-Dimethylpropane	3.11	3.34	-0.23	Pyrazine ^{TS}	-0.26	-0.31	0.05
Cyclopropane ^{TS}	1.72	1.50	0.22	2-Methylpyrazine	0.21	0.36	-0.15

Cyclopentane	3.00	2.61	0.39	Pyrazine-2-carbonitrile	-0.01	-0.30	0.29
Cyclohexane	3.34	2.63	0.71	Ethyl pyrazine-2-carboxylate	0.28	0.66	-0.38
1,2-Dimethylbenzene	3.12	3.30	-0.18	Methyl pyrazine-2-carboxylate ^{TS}	-0.23	-0.22	-0.01
1,3-Dimethylbenzene	3.20	3.14	0.06	3,5-Dimethylpyrazin-2-amine	0.93	1.42	-0.49
1,4-Dimethylbenzene ^{TS}	3.15	3.09	0.06	2-Ethoxypyrazine ^{TS}	1.28	1.09	0.19
1,2,3,4,5,6-hexamethylbenzene	4.61	4.37	0.24	2-Methoxypyrazine	0.73	0.56	0.17
1,2,3,4-Tetrahydronaphthalene	3.49	3.10	0.39	1-Pyrazin-2-ylethanone ^{TS}	0.20	-0.40	0.60
Furan ^{TS}	1.34	1.41	-0.07	2-Ethylpyrazine	0.69	0.99	-0.30
1-Phenylpyrrole	3.08	3.16	-0.08	2-Phenylpyrazine	2.06	1.51	0.55
Anthracene	4.45	4.49	-0.04	Pyrimidine	-0.44	-0.47	0.03
Phenanthrene	4.46	4.09	0.37	2-Methylpyrimidine ^{TS}	-0.05	0.37	-0.42
Benzonitrile	1.56	1.77	-0.21	Pyrimidine-2-carbonitrile ^{TS}	0.08	-0.42	0.50
1-Phenylethanone	1.58	1.59	-0.01	N,N-dimethylpyrimidin-2-amine	1.07	0.94	0.13
Ethyl benzoate	2.64	2.62	0.02	2-Ethoxypyrimidine	0.74	1.41	-0.67
Methyl benzoate ^{TS}	2.12	1.98	0.14	2-Methoxypyrimidine ^{TS}	0.23	0.58	-0.35
N,N-dimethylaniline	2.31	2.25	0.06	4-Methylpyrimidine ^{TS}	-0.07	0.19	-0.26
Nitrobenzene ^{TS}	1.85	2.00	-0.15	Oxirane	0.30	-0.17	0.47
Ethoxybenzene	2.51	2.50	0.01	Pyridine ^{TS}	0.65	0.53	0.12
Methoxybenzene	2.11	2.34	-0.23	Quinoline ^{TS}	2.03	2.69	-0.66
Acetonitrile	-0.34	-0.16	-0.18	Methyl acetate	0.18	0.08	0.10
Propan-2-one	-0.24	0.08	-0.32	Ethyl propanoate	1.21	1.73	-0.52
Methoxymethane	0.10	-0.53	0.63	Pentan-2-one	0.91	0.92	-0.01
Cyclohexanone	0.81	0.79	0.02	Propanenitrile	0.16	0.45	-0.29
1,2-oxazole	0.08	0.33	-0.25	Ethyl-dihydroxyazanium	0.18	0.56	-0.38
Benzaldehyde	1.48	1.39	0.09	4-Ethoxypyrimidine	0.97	1.06	-0.09
Propanal	0.59	0.45	0.14	5-Methylpyrimidine	0.01	0.38	-0.37
2-Phenylacetonitrile ^{TS}	1.56	1.86	-0.30	Pyrimidine-5-carbonitrile	-0.42	-0.34	-0.08
1-phenylpropan-2-one ^{TS}	1.44	1.80	-0.36	Ethyl pyrimidine-5-carboxylate	0.52	0.71	-0.19
N,N-dihydroxymethanamine	-0.35	-0.28	-0.07	Eethyl pyrimidine-5-carboxylate	0.03	-0.12	0.15
2-Phenylacetamide ^{TS}	0.45	1.34	-0.89	5-Ethoxypyrimidine ^{TS}	0.56	1.11	-0.55
2-Phenylacetic acid	1.41	1.91	-0.50	5-Methoxypyrimidine ^{TS}	0.07	0.29	-0.22

Cuadro 9.7 – Valores experimentales, predichos por el modelo y los correspondientes residuos de log P, para el modelo propuesto. Las moléculas pertenecientes al conjunto de prueba está identificadas con el superíndice *TS*.

El poder predictivo del modelo es satisfactorio, esto es revelado por su estabilidad ante la inclusión y o exclusión de compuestos en el conjunto de ajuste, medida por los parámetros estadísticos $R_{loo} = 0,9725$ ($R_{loo}^2 = 0,9458$) y $R_{l-30\%-o} = 0,9553$ ($R_{l-30\%-o}^2 = 0,9125$). Como regla general, $R_{l-n\%-o}(Q)$ debe ser mayor que 0.71 ($Q^2 > 0,5$) para que el modelo sea válido[74, 48]. El modelo también fue validado mediante las siguientes condiciones previamente discutidas[75, 74]: $R_{TS}^2 = 0,8876 > 0,6$; $k = 0,9379 > 0,85$; $k' < 1,15$; $(r^2 - r_0^2)/r^2 = -0,1168 < 0,1$; $(r^2 - r_0'^2)/r^2 = -0,1265 < 0,1$.

La figura 9.1 es una representación de la relación entre los valores experimentales con aquellos predichos empleando el modelo. Puede observarse la relación lineal esperable.

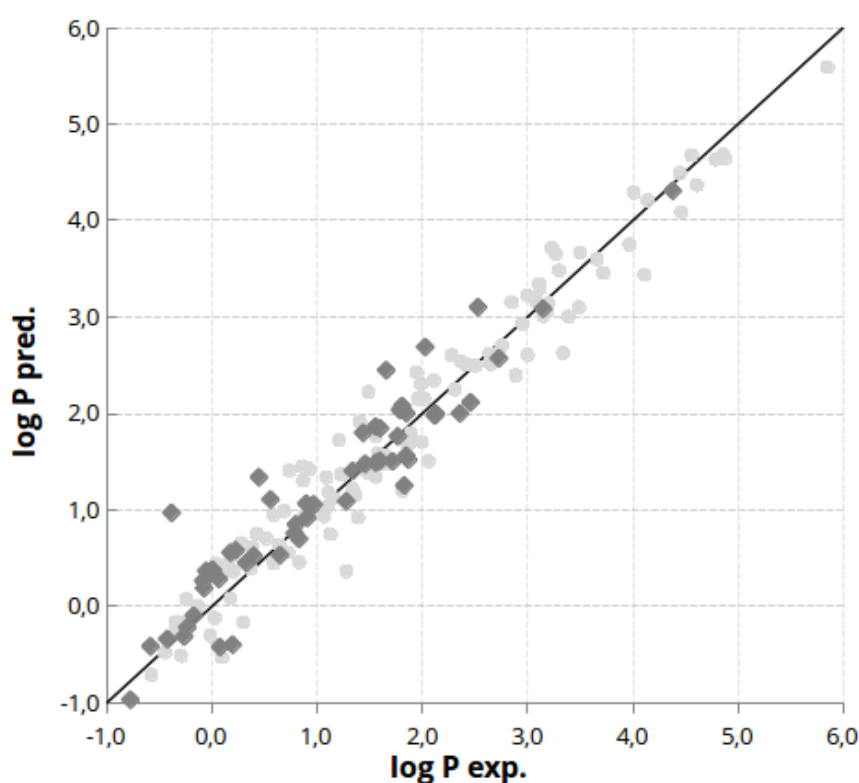


Figura 9.1 – Relación entre los valores de log P predichos y los experimentales. Los valores predichos fueron calculados con la ec. 9.1. Las moléculas pertenecientes al conjunto de ajuste y al conjunto de prueba son representadas con círculos y rombos, respectivamente.

El gráfico de Williams de los residuos estandarizados vs. los valores de los leverages se encuentra en la Figura 9.2. Puede observarse que la mayoría de los compuestos se encuen-

tran en el dominio de aplicabilidad del modelo (ec. 9.1). Los compuestos 40 ($h=0,2913$) y 43 ($h=0,2149$) son h -outliers ($h > 0,1962$) del conjunto de validación. Sus valores de residuos estandarizados son aceptables, lo cual es un indicio de que el modelo es robusto.[50]. El residuo estandarizado del compuesto 95 es 4,2, es un outlier dado que este valor es mayor que el valor límite (3). Este compuesto pertenece al dominio de aplicabilidad del modelo, por lo que la predicción obtenida es un caso atípico. El compuesto 29 excede el valor límite ($h^* = 0,1962$) por un margen considerable. Según se ve en el gráfico 9.2, no se observa una tendencia clara de aumento de los residuos estandarizados con h , para los rangos de h explorados. El valor del compuesto 29 es entonces otro indicio de la confiabilidad del modelo.

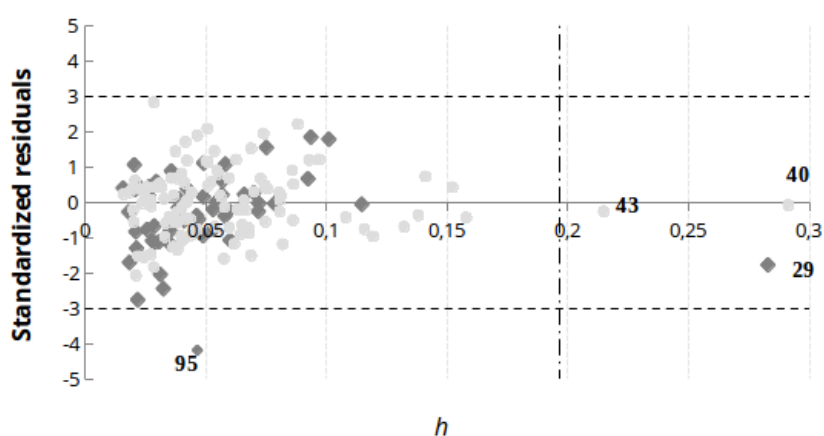


Figura 9.2 – Gráfico de Williams de los residuos estandarizados (obtenidos con la ec. 9.1). Las líneas punteadas verticales indican el leverage límite h^* . Las moléculas pertenecientes al conjunto de ajuste y al conjunto de prueba son representadas con círculos y rombos, respectivamente.

Parte VI

APÉNDICES

Apéndice A

Segunda cuantización

En este apéndice se describirá parte del formalismo de la mecánica cuántica llamado segunda cuantización, ya que se lo utilizará en la descripción del cálculo de propiedades magnéticas. El tratamiento dado aquí está particularizado a sistemas fermiónicos, y restringido a las necesidades de este trabajo de tesis.

La sección A.1 contiene una descripción formal, donde se expone la estructura matemática del formalismo. En la sección A.2 se describe una adaptación al tratamiento tradicionalmente dado en textos de química cuántica, y se resumen los aspectos de dicho tratamiento que aquí interesan.

A.1. Estructura formal

Sean \mathcal{H}_1 y \mathcal{H}_2 dos espacios de Hilbert separables, definidos sobre el cuerpo complejo \mathbb{C} . Para cada $\psi \in \mathcal{H}_1$ y cada $\phi \in \mathcal{H}_2$, puede definirse una forma bilineal $\psi \otimes \phi : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ con

$$[\psi \otimes \phi](x, y) = \langle x, \psi \rangle \langle y, \phi \rangle$$

y $(x, y) \in \mathcal{H}_1 \times \mathcal{H}_2$.

Esta definición junto con las definiciones usuales de suma de vectores y producto por un escalar, son suficientes para definir un espacio tensorial \mathcal{E} . Defínase un producto interior sobre este espacio según

$$\langle \psi_1 \otimes \phi_1, \psi_2 \otimes \phi_2 \rangle = \langle \psi_1, \psi_2 \rangle \langle \phi_1, \phi_2 \rangle$$

con $\psi_1, \psi_2 \in \mathcal{H}_1$ y $\phi_1, \phi_2 \in \mathcal{H}_2$. Cualquiera de los espacios resultantes de completar \mathcal{E} ,

respecto de la métrica obtenida a partir del producto interior, será llamado *producto tensorial* de \mathcal{H}_1 y \mathcal{H}_2 , y será denotado por $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Comentario A.1.1. Si (M, d) es un espacio métrico, se llama *completación* de (M, d) a $(f, (\tilde{M}, \tilde{d}))$ donde (\tilde{M}, \tilde{d}) es un espacio métrico completo, y $f : (M, d) \rightarrow (\tilde{M}, \tilde{d})$ es una isometría tal que $f(M)$ es denso en \tilde{M} . Dos completaciones de un dado espacio métrico son necesariamente isométricamente isomorfas. A partir de todo espacio métrico puede obtenerse un espacio completo.

Se define al **espacio de Fock** \mathcal{F} con $\mathcal{F}(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathcal{H}^{\otimes n}$.
y al **operador de antisimetrización** de n partículas, $\mathcal{S}_-^{(n)}$,

$$\mathcal{S}_-^{(n)}(f_1 \otimes \cdots \otimes f_n) := \frac{1}{n!} \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(n)}$$

$$\mathcal{S}_-^{(0)} := \mathbf{1}$$

Siendo S_n un grupo de permutaciones σ , que mapean según

$$\sigma(v_1, \dots, v_n) = (v_{\sigma(1)}, \dots, v_{\sigma(n)})$$

$\text{sgn}:=1$ si la permutación es par, y $\text{sng}:-1$ si es impar¹.

Se llama **espacio de Fock antisimétrico** o **espacio de Fock fermiónico**, sobre \mathcal{H} , al espacio $\mathcal{F}_-(\mathcal{H}) := \bigoplus_{n=0}^{\infty} \mathcal{S}_-^{(n)}(\mathcal{H}^{\otimes n})$.

Al vector $\mathbf{1} \in \mathcal{S}_-^{(0)}(\mathcal{H}^{\otimes 0})$ se lo llama **estado vacío**, $|\Omega\rangle$. Se define a un **operador creación**, a^\dagger , para cada $f \in \mathcal{H}$, según

$$a^\dagger(f) \mathcal{S}_-^{(n)}(f_1 \otimes \cdots \otimes f_n) := \sqrt{n+1} \mathcal{S}_-^{(n+1)}(f \otimes f_1 \otimes \cdots \otimes f_n) \quad (\text{A.1})$$

Análogamente se define al **operador aniquilación** con

$$a(f) \mathcal{S}_-^{(n)}(f_1 \otimes \cdots \otimes f_n) := \frac{1}{\sqrt{n}} \sum_j (-1)^{(j-1)} \langle f, f_j \rangle \mathcal{S}_-^{(n-1)}(f \otimes f_1 \otimes \cdots \otimes f_n)$$

$$a(|\Omega\rangle) := |0\rangle \quad (\text{A.2})$$

Dada una base ortonormal de \mathcal{H} , $\{\mathbf{e}_i\}$, puede generarse una base ortonormal de $\mathcal{S}_-^{(n)}(\mathcal{H}^{\otimes n})$, a partir de vectores formados según

$$\mathbf{e}(n_1, n_2, \dots) := \sqrt{n!} \mathcal{S}_-^{(n)}(\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_n}) \quad (\text{A.3})$$

¹Una permutación es par si puede descomponerse en un número par de trasposiciones (entre dos elementos), si una transposición no es par entonces es impar.

donde $i_1 < i_2 < \dots < i_n$. $n_i = 1$ en caso que \mathbf{e}_i se encuentre entre $\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_n}$, y $n_i = 0$ en caso contrario. Se llama a $n = \sum_i n_i$ **número de ocupación**.

Se definen los operadores $a_i^\dagger := a^\dagger(\mathbf{e}_i)$ y $a_i := a(\mathbf{e}_i)$, que según las ecuaciones A.1, A.2 y A.3

$$a_i^\dagger e(n_1, \dots, n_i, \dots) = \begin{cases} (-1)^{s_i} \mathbf{e}(n_1, \dots, n_i + 1, \dots) & \text{si } n_i = 0 \\ 0 & \text{si } n_i = 1 \end{cases}$$

$$a_i e(n_1, \dots, n_i, \dots) = \begin{cases} 0 & \text{si } n_i = 0 \\ (-1)^{s_i} \mathbf{e}(n_1, \dots, n_i - 1, \dots) & \text{si } n_i = 1 \end{cases}$$

siendo s_i el número de índices i_j que satisfacen $i_j < i$.

Dado un operador \mathcal{A} auto-adjunto que opere sobre \mathcal{H} , puede definirse un operador $d\Gamma(A) := \sum_{n=0}^{\infty} \sum_{v=1}^n \mathbf{1} \otimes \dots \otimes \mathcal{A} \otimes \dots \otimes \mathbf{1} = \sum_{i,j} A_{ij} a_i^\dagger a_j$, siendo $A_{ij} := \langle \mathbf{e}_i, A(\mathbf{e}_j) \rangle$.

De esta forma se obtiene un nuevo operador que actúa de la misma forma sobre cada partícula individualmente.

Los operadores $a^\dagger(f)$ y $a(f)$ son acotados y sus normas cumplen con $\|a^\dagger(f)\| = \|a(f)\| = \|f\|$. El operador $a^\dagger(f)$ es lineal y el operador $a(f)$ es antilineal, y son mutuamente adjuntos.

Puede demostrarse que se cumplen las *relaciones de anticonmutación canónicas*:

$$\langle f, g \rangle = [a(f), a^\dagger(g)]_+ \\ 0 = [a(f), a(g)]_+ = [a^\dagger(f), a^\dagger(g)]_+$$

A.2. Aplicación en química computacional

En el marco de la aproximación de Born-Oppenheimer, una molécula puede ser tratada aproximadamente como un sistema n -electrónico, inmerso en focos de potenciales puntuales y fijos, dados por los núcleos atómicos. Dicho sistema tiene solución exacta dada por una combinación lineal de determinantes de Slater

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \dots & \phi_n(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \dots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_n) & \phi_2(\mathbf{x}_n) & \dots & \phi_n(\mathbf{x}_n) \end{vmatrix} \equiv |\phi_1 \ \phi_2 \ \dots \ \phi_n|$$

de modo que el conjunto de dichos determinantes es un conjunto de generadores del espacio de las soluciones. En la ecuación anterior las $\phi_i(\cdot)$ representan funciones de las variables de espín y espacio de del electrón j , (\mathbf{x}_j).

Supóngase la existencia de un sistema de n partículas desprovistas de espín. Su vector de estado será $|\Psi(\mathbf{r}_1, \dots, \mathbf{r}_n)\rangle$, y podrá ser representado por una función de las coordenadas espaciales $\{\mathbf{r}_i\}$. En caso que dicha función pueda expresarse como el producto de funciones monoelectrónicas $\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\dots\phi_n(\mathbf{r}_n)$, el vector de estado puede representarse con

$$|\Phi_n\rangle = |\phi_1\rangle \otimes |\phi_2\rangle \otimes \dots \otimes |\phi_n\rangle$$

Donde $|\Phi_n\rangle \in \mathcal{H}^{\otimes n}$, y $|\phi_i\rangle \in \mathcal{H}^i(2)$. Si $\{|u_i^{(k)}\rangle\}$ es una base de \mathcal{H}^k (del espacio de la partícula k), entonces los vectores $\{|u_i^{(1)}\rangle \otimes \dots \otimes |u_j^{(n)}\rangle\}$ son base de $\mathcal{H}^{\otimes n}$.

Dado que los electrones son fermiones, no todos los los vectores y funciones en $\{\mathcal{H}^{\otimes n}\}(n > 1)$ recién descriptos satisfacen el postulado de simetría, entonces los espacios $\{\mathcal{H}^{\otimes n}\}(n > 1)$, contienen funciones no válidas. Es posible evitarlas usando espacios $\{\mathcal{H}_-^{\otimes n}\}(n > 0)$ generados por el mapeo de $\mathcal{S}_-^{(n)}$ sobre los $\{\mathcal{H}^{\otimes n}\}(n > 0)$. Desde un punto de vista práctico, para cada espacio (\mathcal{H}^i) se usa bases con funciones idénticas, que consisten en un conjunto ortonormal de espín-orbitales $\{\phi_p\}$. Los productos de Hartree serían vectores de $\mathcal{H}^{\otimes n}$ y la antisimetrización se logra por restricción a determinantes de Slater, pertenecientes a $\mathcal{H}_-^{\otimes n}$.

Dado que en la práctica no puede usarse una base de infinitos vectores, se restringe el espacio a aquel generado por M espín-orbitales. En este contexto, la ecuación A.3 puede ser reemplazada por

$$\mathbf{e}(n_1, \dots, n_M) := \sqrt{n!} \mathcal{S}_-^{(n)}(\phi_{p_1} \otimes \dots \otimes \phi_{p_n})$$

donde $\mathbf{e}(n_1, \dots, n_M)$ corresponde ahora a un determinante de Slater. Si un dado ϕ_p , se encuentra entre los $\phi_{p_1}, \dots, \phi_{p_n}$, ϕ_p se dice de él que está ocupado, en caso contrario que está desocupado. Dicho determinante corresponde al **vector número de ocupación**,

$$|\mathbf{k}\rangle := |k_1, \dots, k_M\rangle \wedge k_p = \begin{cases} 1 & \text{si } \phi_p \text{ está ocupado} \\ 0 & \text{si } \phi_p \text{ está desocupado} \end{cases}$$

A partir de un conjunto de espín orbitales puede obtenerse otro mediante la aplicación de una transformación unitaria

$$\tilde{\phi}_p := \sum_q \phi_q U_{qp}$$

²Naturalmente $\phi_i(\mathbf{r}_i)$ es la función que representa al vector abstracto $|\phi_i\rangle \in \mathcal{H}^i$.

y la matriz unitaria U puede ser escrita en la forma exponencial

$$U = \exp(-\kappa)$$

donde κ es una matriz antihermítica ($\kappa^\dagger = -\kappa$).

Esta matriz puede expresarse en términos de los operadores de creación y aniquilación vinculados a los orbitales originales

$$\sum_{pq} \kappa_{pq} a_p^\dagger a_q$$

En esta expresión κ_{pq} son los elementos de matriz de κ . Esta matriz permite vincular al espacio de Fock asociado a los primeros orbitales, con un segundo espacio de Fock asociado a los nuevos orbitales mediante las transformaciones. Las transformaciones correspondientes para los operadores de creación y aniquilación, y algún estado cualquiera $|0\rangle$ son

$$\begin{aligned}\tilde{a}_p^\dagger &= \exp(-\kappa) a_p^\dagger \exp(\kappa) \\ \tilde{a}_p &= \exp(-\kappa) a_p \exp(\kappa) \\ |\tilde{0}\rangle &= \exp(-\kappa) |0\rangle\end{aligned}$$

Estos operadores permiten realizar rotaciones generales. Puede preferirse el uso de operadores que generen un conjunto restringido de rotaciones. Una posibilidad es cuando se trabaja con orbitales reales y no se necesite más que rotaciones reales. En el cálculo de las propiedades magnéticas el uso de rotaciones imaginarias es necesario. Una segunda posibilidad es la restricción por grupo puntual de simetría. En tal caso sólo rotaciones totalmente simétrica deben ser empleadas para preservar la simetría del vector de estado.

El tercer caso es rotaciones adaptadas por espín. Como el operador hamiltoniano y el de espín conmutan, suele preferirse trabajar con rotaciones que preserven el espín. En tales casos se construye al operador de rotación en términos de los operadores de excitación singlete E_{pq} y triplete T_{pq}

$$\begin{aligned}
E_{pq} &:= a_{p\alpha}^\dagger a_{q\alpha} + a_{p\beta}^\dagger a_{q\beta} \\
\mathbf{T}_{pq} &:= \{T_{pq}^x, T_{pq}^y, T_{pq}^z\} \\
T_{pq}^x &:= \frac{1}{2}(a_{p\alpha}^\dagger a_{q\beta} + a_{p\beta}^\dagger a_{q\alpha}) \\
T_{pq}^y &:= \frac{1}{2i}(a_{p\alpha}^\dagger a_{q\beta} - a_{p\beta}^\dagger a_{q\alpha}) \\
T_{pq}^z &:= \frac{1}{2}(a_{p\alpha}^\dagger a_{q\alpha} - a_{p\beta}^\dagger a_{q\beta})
\end{aligned}$$

La parte real e imaginaria de las rotaciones pueden separarse si se reescribe por conveniencia al operador de rotación como

$$\boldsymbol{\kappa} = i \sum_p {}^I\kappa_{pp} a_p^\dagger a_p + i \sum_{p>q} {}^I\kappa_{pq} (a_p^\dagger a_q + a_q^\dagger a_p) + \sum_{p>q} {}^R\kappa_{pq} (a_p^\dagger a_q - a_q^\dagger a_p)$$

con

$$\begin{aligned}
{}^R\kappa_{pq} &:= \text{Re}(\kappa_{pq}) = \frac{\kappa_{pq} - \kappa_{qp}}{2} \\
{}^I\kappa_{pq} &:= \text{Im}(\kappa_{pq}) = \frac{\kappa_{pq} + \kappa_{qp}}{2i}
\end{aligned}$$

Finalmente definiendo

$$\begin{aligned}
E_{pq}^\pm &:= E_{pq} \pm E_{qp} \\
T_{pq}^{\alpha\pm} &:= T_{pq}^\alpha \pm T_{qp}^\alpha
\end{aligned}$$

se obtiene

$$\boldsymbol{\kappa} = i \sum_{p>q} {}^I\kappa_{pq}^0 + \sum_{p>q} {}^R\kappa_{pq}^0 E_{pq}^- + i \sum_\alpha \sum_{p>q} {}^I\kappa_{pq}^\alpha T_{pq}^{\alpha+} + \sum_\alpha \sum_{p>q} {}^R\kappa_{pq}^\alpha T_{pq}^{\alpha-} \quad (\text{A.4})$$

Apéndice B

Códigos de fuente

En este apéndice se reseña los códigos de fuente con mayor importancia esta tesis. Se escogió el lenguaje Python 3 por su simpleza. Se priorizó la sencillez y velocidad de creación del código por sobre la eficiencia. Para los procesos más intensivos computacionalmente se utilizó la biblioteca Numpy[76] y otras derivadas de esta.

Cálculo de los descriptores

En esta sección se describe los códigos desarrollados para el cálculo de los descriptores. Los mismos utilizan información de archivos de salida de Gaussian09, que deben contener la información de los cálculos de las propiedades magnéticas. Los últimos deben ser ubicados en un directorio en el cual se incluya al siguiente programa escrito en Python3. El código del programa está dividido en varios archivos. El siguiente código debe ser ubicado en la raíz del directorio y ejecutado por el interprete de Python.

```
1 import numpy as np
2 import os, sys
3
4 sys.path.append(os.getcwd()+"/lib")
5
6 from elements.elements import ELEMENTS as elemento
7 from gaussian import g0x
8 from gan import gan
9 from descriptores import descriptores as qspr
10
11 archivos = [x for x in os.listdir(os.getcwd()) if x.endswith(".log")]
12 from natsort import natsorted as natural_sort
13 archivos = natural_sort(archivos, reverse=True)
14
15 get_prop = {
```

```

16     "fc_j" : g0x.fc_j,
17     "fc_k" : g0x.fc_k,
18     "sd_j" : g0x.sd_j,
19     "sd_k" : g0x.sd_k,
20     "pso_j" : g0x.pso_j,
21     "pso_k" : g0x.pso_k,
22     "dso_j" : g0x.dso_j,
23     "dso_k" : g0x.dso_k,
24     "tnss_j": g0x.tnss_j,
25     "fnss_k": g0x.tnss_k,
26     "shield": g0x.shield_iso_ev
27 }
28
29 prop_list = sorted( list(get_prop.keys()) )
30
31 descript_set = sorted( list(qspr.calc_descriptor.keys()) )
32
33 name_list = []
34
35 inp = open(archivos[0], "r")
36 f = inp.readlines()
37 inp.close()
38 a = get_prop["shield"](f)
39 a = qspr.transform(a)
40 mol = g0x.geom(f)
41 dt = gan.dist_t_mat(mol)
42 rij = gan.rij_mat(mol)
43 xyz = gan.only_xyz(mol)
44
45 largo = []
46 for i in descript_set:
47     if ( i == "getaway" ):
48         largo.append( len(qspr.calc_descriptor[i](rij,a,dt,xyz)) )
49     elif ( i == "morse3d" ):
50         largo.append(32)
51     elif ( i == "rdf" ):
52         largo.append( len( qspr.calc_descriptor[i](rij,a) ) )
53     elif ( i == "whim" ):
54         largo.append(5)
55     elif ( i == "ats" ):
56         largo.append(4)
57
58 for i in prop_list:
59     for idx, j in enumerate(descript_set):
60         for k in range(0,largo[idx]):
61             name_list.append (i+"-"+j+"-"+str(k))
62 argumentos = {
63     "getaway" : (rij,a,dt,xyz),
64     "morse3d" : (rij,a),
65     "rdf"      : (rij,a),
66     "whim"     : (a, xyz),
67     "ats"      : (a, dt)
68 }
69

```

```

70 descriptors = np.empty((len(archivos),len(name_list)))
71
72 for idx, i in enumerate(archivos):
73     inp = open(i, "r")
74     f = inp.readlines()
75     inp.close()
76     fila = []
77     print("Molécula: ", i)
78     mol = g0x.geom(f)
79     dt = gan.dist_t_mat(mol)
80     rij = gan.rij_mat(mol)
81     xyz = gan.only_xyz(mol)
82     for j in prop_list:
83         print("... ", j)
84         a = get_prop[j](f)
85         a = qspr.transform(a)
86         argumentos = {
87             "getaway" : (rij,a,dt,xyz),
88             "morse3d" : (rij,a),
89             "rdf"      : (rij,a),
90             "whim"     : (a, xyz),
91             "ats"      : (a, dt)
92         }
93         for k in descript_set:
94             print(".....", k)
95             fila.append( qspr.calc_descriptor[k](*argumentos[k]) )
96     fila = [val for sublist in fila for val in sublist]
97     descriptors[idx] = np.array( fila )
98
99 no_numericos = []
100 for i in range(0,descriptors.shape[1]):
101     if (np.isnan(descriptors[:,i]).any()):
102         no_numericos.append(i)
103
104 descriptors = np.delete( descriptors , no_numericos, axis=1)
105
106 np.savetxt('descriptors_values.txt', descriptors )
107
108 name_list = [i for j, i in enumerate(name_list) if j not in no_numericos]
109
110 np.savetxt('descriptors_names.txt', np.array(name_list), fmt="%s" )

```

Este código es la interfaz principal del programa. Su principal función es delegar la tarea de analizar los archivos de salida y el calcular los descriptores, a módulos descritos a continuación. También, verificar y escribir archivos con la información final.

El resto de los módulos deberían ubicarse en un subdirectorio con nombre `lib`. Se da como ejemplo el arbol de directorio utilizado. El mismo es el correspondiente a antes de la ejecución del programa. Luego de la ejecución los archivos de salida estarían presentes¹. Los

¹Y los archivos de código intermedio “.pyc”.

archivos `.log` corresponden a las salidas de Gaussian y `run_me.py` al fichero principal.

```

├─ db162-0.log
├─ db162-1.log
├─ ...
├─ db162-161.log
├─ run_me.py
└─ lib
    ├─ descriptores
    │   └─ descriptores.py
    ├─ elements
    │   └─ elements.py
    ├─ gan
    │   └─ gan.py
    └─ gaussian
        ├─ g0x.py
        └─ __init__.py

```

El fichero `descriptores.py` contiene las funciones de creación de los descriptores, y su código se encuentra al finalizar este párrafo. Las mismas realizan muy poco análisis de la estructura de la molécula.

```

1 import numpy as np
2
3 def transform(x):
4     if (type(x) is dict):
5         l = int( np.sqrt( 2*len(x) ))
6         a = np.empty([l, l])
7         for key in x:
8             a[key[0]-1,key[1]-1] = x[key]
9             a[key[1]-1,key[0]-1] = x[key]
10        return(a)
11    a = np.array(x)
12    if (hasattr(a[0], "__len__") ):
13        return(a)
14    else:
15        return(np.outer(a,a))
16    return(a)
17
18 def morse3d(rij, a, **kwargs):
19     lista = np.linspace(0.,31.,32)
20     for i in range(len(rij[0])-1):
21         for j in range(i+1,len(rij[0]) ):
22             lista[0] += a[i][j]
23     for k in range(1,32):

```

```

24     for i in range(len(rij[0])-1):
25         for j in range(i+1, len(rij[0])):
26             lista[k] += a[i][j]*np.sin( rij[i][j]*float(k) )/ ( rij[i][j]*float(k) )
27     return(lista)
28
29 def rdf(rij, a, **kwargs):
30     l = np.linspace(0., 10., 60).tolist()
31     lista = np.array([0.]*60)
32     if ('beta') in kwargs:
33         beta = kwargs["beta"]
34     else:
35         beta = 150.
36     for k, m in enumerate(l):
37         for i in range(len(rij[0])-1):
38             for j in range(i+1, len(rij[0])):
39                 lista[k] += a[i][j]*np.exp( -beta*(rij[i][j]-float(m))**2 )
40     return(lista)
41
42 def getaway(rij, a , dt, M):
43     M = np.array(M)
44     mean = np.empty(3)
45     for i in [0, 1, 2]:
46         mean[i] = np.mean(M[:,i])
47     for i in range(0, len(M)):
48         for j in [0,1,2]:
49             M[i,j] = M[i,j] - mean[j]
50
51     H = (M.dot( np.linalg.pinv( (M.T).dot(M) ) )).dot(M.T)
52     R = np.copy(H)
53
54     for i in range(0, len(R[0])):
55         for j in range(i, len(R[0])):
56             if ( i != j):
57                 R[i][j] = np.sqrt(H[i][i]*H[j][j])/rij[i][j]
58                 R[j][i] = R[i][j]
59             else:
60                 R[i][i] = 0.
61
62     HATS_k = []
63     k_max = np.amax(dt)
64     for k in range(0, k_max):
65         HATS_k.append(0.)
66         for i in range(0, len(R[0])):
67             for j in range(0, len(R[0])):
68                 if (k == dt[i][j]):
69                     HATS_k[k] += a[i][j]*H[i][i]*H[j][j]
70     HATS = HATS_k[0]
71     for i in range(1, len(HATS_k)):
72         HATS += 2. * HATS_k[i]
73
74     H_k = []
75     for k in range(0, k_max):
76         H_k.append(0.)
77         for i in range(0, len(R[0])):

```

```

78         for j in range(0, len(R[0])):
79             if (k == dt[i][j] and H[i][j] > 0.):
80                 H_k[k] += a[i][j]*H[i][j]
81     HT = H_k[0]
82     for i in range(1,len(H_k)):
83         HT += 2. * H_k[i]
84
85     R_k = [None]
86     for k in range(1, k_max):
87         R_k.append(0.)
88         for i in range(0,len(R[0])):
89             for j in range(0, len(R[0])):
90                 if (k == dt[i][j]):
91                     R_k[k] += a[i][j]*R[i][j]
92     RT = 0.
93     for i in range(1,len(R_k)):
94         RT += R_k[i]
95     RT = 2. * RT
96
97     return(HATS,HT,RT)
98
99
100 def whim(a, M):
101     M = np.array(M)
102     S = np.zeros((3,3))
103     M = M.T
104     mean = np.empty(3)
105
106     mean[0] = np.mean(M[0])
107     mean[1] = np.mean(M[1])
108     mean[2] = np.mean(M[2])
109
110     A = np.trace(a)
111     if ( A == 0.):
112         return(None, None, None, None, None)
113
114     for i in [0, 1, 2]:
115         for j in range(i,3):
116             for k in range(0, len(S)):
117                 S[i][j] += a[k][k] * ( M[i][k] - mean[i] ) * ( M[j][k] - mean[j] )
118                 S[i][j] = S[i][j] / A
119                 S[j][i] = S[i][j]
120
121     av = sorted ( np.linalg.eigvals(S) )
122
123     T = np.sum( av )
124     s = av[0]*av[1] + av[0]*av[2] + av[2]*av[1]
125
126     return(T, s, av[0]/T, av[1]/T, av[2]/T)
127
128
129 def ats(a , dt):
130     atskw = [0.,0.,0.,0.]
131     for k in range(0, 4):

```



```

132     for i in range(len(a[0])-1):
133         for j in range(i,len(a[0])):
134             if (k == dt[i][j]):
135                 atskw[k] += a[i][j]
136     return(atskw)
137
138
139 calc_descriptor={
140     "morse3d": morse3d,
141     "rdf" : rdf,
142     "getaway": getaway,
143     "whim" : whim,
144     "ats" : ats
145 }

```

El análisis estructural básico, que incluye la obtención de matrices de distancias interatómicas, topológicas², etc., fue realizado con funciones pertenecientes a una biblioteca para análisis estructural creada por quien escribe. A continuación se encuentra el código de las funciones utilizadas

```

1 import numpy as np
2 from elements.elements import ELEMENTS as elemento
3 import networkx as nx
4
5 def d(p1, p2):
6     p = np.array(p1) - np.array(p2)
7     return(np.asscalar(np.sqrt((p).dot(p.T))))
8
9 def only_xyz(molec):
10     natom = len(molec)
11     xyz = []
12     for i in range(natom):
13         xyz.append(
14             [float(molec[i][1]), float(molec[i][2]), float(molec[i][3])])
15     return(xyz)
16
17 def rij_mat(mol):
18     mol = np.array(only_xyz(mol))
19     # Crea la matriz de distancias interatómicas
20     rij = np.empty([len(mol), len(mol)])
21     for i in range(len(mol)):
22         for j in range(i,len(mol)):
23             rij[i][j] = np.linalg.norm(mol[i]-mol[j])
24             rij[j][i] = rij[i][j]
25     return(rij)
26
27 def conn_atom(at1, at2, **kwargs):
28     if ('tol' in kwargs):
29         tol = float(kwargs['tol'])/100.

```

²En este contexto suele llamarse así al mínimo número necesario de enlaces covalentes a recorrer para conectar dos átomos.

```

30     else:
31         tol = 0.2
32         dc = elemento[at1[0]].covrad + elemento[at2[0]].covrad
33         dr = d( only_xyz([at1]) , only_xyz([at2]) )
34         if (dr <= dc *(1.+tol)):
35             return(True)
36     else:
37         return(False)
38
39 def conn_mol_list(mol, **kwargs):
40     if ('tol' in kwargs):
41         tol = kwargs['tol']
42     else:
43         tol = 20.
44
45     dic = dict()
46     for i in range(0,len(mol)):
47         dic[i] = []
48     for i in range(0,len(mol)-1):
49         for j in range(i+1,len(mol)):
50             if conn_atom(mol[i],mol[j],tol=tol):
51                 dic[i].append(j)
52                 dic[j].append(i)
53     return(dic)
54
55 def dist_t_mat(mol):
56     graph = conn_mol_list(mol)
57     x = nx.from_dict_of_lists(graph)
58     x = nx.all_pairs_shortest_path_length(x)
59     sh = np.empty([len(mol), len(mol)], dtype=np.int)
60     for i in range(len(mol)):
61         for j in range(i,len(mol)):
62             sh[i,j] = int(x[i][j])
63             sh[j,i] = sh[i,j]
64     return(sh)

```

La recuperación de la información de los archivos de salida de Gaussian 09 se comentará con mayor detalle utilizando un ejemplo. Se usó la línea de palabras claves comentada en el texto³

```
#p B3LYP/6-311+G(2d,p) NMR(GIA0,PrintEigenvectors,Susceptibility,SpinSpin)
SCRF=(Solvent=Water)
```

para la molécula neutra de geometría (coordenadas cartesianas y Armstrong)

C	0.6414	-1.2059	0.0001
C	-0.1656	-0.0586	0.
N	1.9758	-1.1352	0.0001
N	0.365	1.1737	-0.0002
C	-1.6668	-0.1941	-0.0002

³La línea fue dividida aquí por cuestiones de espacio.

C	1.6982	1.2445	-0.0001
C	2.4972	0.0977	0.
O	-2.1719	-1.3045	-0.0002
C	-2.4855	1.0756	0.0001
H	0.1816	-2.1895	0.0001
H	2.1444	2.2365	0.
H	3.5821	0.1761	0.
H	-2.2411	1.6883	-0.8731
H	-3.5442	0.8158	-0.0028
H	-2.2456	1.6842	0.8776

Los tensores de apantallamiento magnético son impresos secuencialmente. El correspondiente al primer átomo verse como

```
Calculating GIAO nuclear magnetic shielding tensors.
SCF GIAO Magnetic shielding tensor (ppm):
1 C   Isotropic =    31.6738   Anisotropy =   181.8747
XX=   -1.1885   YX=   -42.0855   ZX=   -0.0031
XY=   -24.7047   YY=   -56.7137   ZY=    0.0087
XZ=    0.0005   YZ=    0.0126   ZZ=   152.9236
Eigenvalues:   -72.3791    14.4769   152.9236
Eigenvectors:
(1)   0.424689   0.905339  -0.000040
(2)   0.905339  -0.424689   0.000041
(3)  -0.000020   0.000054   1.000000
...
```

en donde se agregó colores para resaltar valores. En el caso de las propiedades derivadas del tensor de apantallamiento, de acuerdo con la descripción del texto principal, se usó la componente isotrópica para asignar valores a los símbolos A_k de las formas funcionales de los descriptores. En este ejemplo: $A_1 = 31,6738$.

En las salidas de Gaussian, las diferentes contribuciones a los tensores de acoplamiento indirecto (**J** y **K**) son impresas en la siguiente forma. Como ejemplo se muestra el aporte a **K** del contacto de Fermi en la molécula de ejemplo

```
Fermi Contact (FC) contribution to K (Hz):
1           2           3           4           5
1  0.000000D+00
2  0.332974D+02  0.000000D+00
3  0.390108D+01 -0.269880D+01  0.000000D+00
4 -0.230553D+01  0.495064D+01 -0.203286D+00  0.000000D+00
5  0.260606D+01  0.308538D+02 -0.564188D-01  0.127820D+02  0.000000D+00
6  0.661404D+01 -0.240181D+01 -0.254889D+01  0.507439D+01  0.284381D+01
```

```

7 -0.269988D+01  0.669031D+01  0.353706D+01 -0.229293D+01 -0.524521D+00
8  0.155884D+01  0.257860D+01  0.241546D+00  0.851439D+00 -0.151657D+02
9  0.274583D+00  0.616291D+01 -0.128164D+00  0.468042D+00  0.223944D+02
10  0.225937D+02  0.105485D+01  0.374914D+01  0.167202D+00  0.881711D-01
11 -0.182781D+00  0.118215D+01  0.267146D+00  0.371187D+01 -0.478286D-01
12  0.129280D+01 -0.200929D+00  0.366512D+01  0.306907D+00  0.139543D+00
13 -0.150646D-01 -0.531560D-01  0.572437D-02  0.199456D+00 -0.688867D+00
14  0.112511D+00  0.480465D+00  0.209099D-01  0.462928D+00 -0.618933D+00
15 -0.152864D-01 -0.533295D-01  0.565885D-02  0.197548D+00 -0.692076D+00
6          7          8          9          10
6  0.000000D+00
7  0.308114D+02  0.000000D+00
8  0.282590D-01  0.252611D+00  0.000000D+00
9 -0.472563D-01  0.184252D+00  0.238535D+01  0.000000D+00
10 -0.176902D+00  0.125898D+01  0.618030D-01 -0.268989D-01  0.000000D+00
11  0.223777D+02  0.147223D+01  0.206077D-01  0.322053D-02  0.437723D-01
12  0.146455D+01  0.222045D+02 -0.341489D-01  0.431217D-03 -0.202792D-01
13  0.114412D-01  0.430523D-02  0.574106D+00  0.153029D+02 -0.113742D-04
14  0.150984D-01  0.185199D-01  0.537229D+00  0.160000D+02  0.333220D-02
15  0.112909D-01  0.425931D-02  0.570540D+00  0.152898D+02 -0.244740D-04
11          12          13          14          15
11  0.000000D+00
12  0.756549D-01  0.000000D+00
13  0.784092D-03  0.268709D-03  0.000000D+00
14  0.443598D-03  0.350386D-03 -0.365634D+00  0.000000D+00
15  0.794009D-03  0.271184D-03 -0.637639D+00 -0.368105D+00  0.000000D+00

```

A partir de esto, según se explicó en el texto, se obtienen los productos $A_i \cdot A_j$, por ejemplo, para $i, j = 4, 7$: $A_4 \cdot A_7 = -2,29293$.

El código que realiza la extracción de esta información, g0x.py en el ejemplo, se encuentra a continuación.

```

1 from elements.elements import ELEMENTS as elemento
2
3 def natoms(file):
4     for line in file:
5         if " NAtoms= " in line:
6             return(int(line.split()[1]))
7
8 def geom(file):
9     lista = []
10    na = natoms(file)
11    for n,line in enumerate(file):
12        if " Standard orientation:" in line:
13            i = n+5
14    while not (len(file[i].split()) < 6):

```

```
15         if (len(file[i].split()) == 6):
16             value = [elemento[int(file[i].split()[1]).symbol,float(file[i].split()[3]),
17                           float(file[i].split()[4]),float(file[i].split()[5])]]
18             lista.append(value)
19             i += 1
20         return (lista)
21
22 ##### INFORMACIÓN DE CÁLCULOS NMR
23 def shield_iso(file):
24     lista = []
25     na = natoms(file)
26     for n,line in enumerate(file):
27         if "SCF GIAO Magnetic shielding tensor" in line:
28             start = n
29     for i in range(0, na):
30         lista.append( float(file[start+1+5*i].split()[4] ) )
31     return (lista)
32
33 def shield_iso_ev(file):
34     lista = []
35     na = natoms(file)
36     for n,line in enumerate(file):
37         if "SCF GIAO Magnetic shielding tensor" in line:
38             start = n
39     for i in range(0, na):
40         lista.append( float(file[start+1+9*i].split()[4] ) )
41     return (lista)
42
43 def shield_aniso(file):
44     lista = []
45     na = natoms(file)
46     for n,line in enumerate(file):
47         if "SCF GIAO Magnetic shielding tensor" in line:
48             start = n
49     for i in range(0, na):
50         lista.append( float(file[start+1+5*i].split()[7] ) )
51     return (lista)
52
53 def fc_k(file):
54     dic = dict()
55     na = natoms(file)
56     for n,line in enumerate(file):
57         if "Fermi Contact (FC) contribution to K (Hz):" in line:
58             start = n
59     n = start + 1
60     resto = na%5
61     parts = int(na/5)
62     if (resto > 0):
63         parts += 1
64     for i in range(0, parts):
65         n = n + 1
66         for j in range(1,na - 5*i+1):
67
68             for k in range (0, min(j,5)):
```

```

69         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
70         n +=1
71     return (dic)
72
73 def fc_j(file):
74     dic = dict()
75     na = natoms(file)
76     for n,line in enumerate(file):
77         if "Fermi Contact (FC) contribution to J (Hz):" in line:
78             start = n
79             n = start + 1
80             resto = na%5
81             parts = int(na/5)
82             if (resto > 0):
83                 parts += 1
84             for i in range(0, parts):
85                 n = n + 1
86                 for j in range(1,na - 5*i+1):
87
88                     for k in range (0, min(j,5)):
89                         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
90                         n +=1
91             return (dic)
92
93 def sd_k(file):
94     dic = dict()
95     na = natoms(file)
96     for n,line in enumerate(file):
97         if "Spin-dipolar (SD) contribution to K (Hz):" in line:
98             start = n
99             n = start + 1
100             resto = na%5
101             parts = int(na/5)
102             if (resto > 0):
103                 parts += 1
104             for i in range(0, parts):
105                 n = n + 1
106                 for j in range(1,na - 5*i+1):
107
108                     for k in range (0, min(j,5)):
109                         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
110                         n +=1
111             return (dic)
112
113 def sd_j(file):
114     dic = dict()
115     na = natoms(file)
116     for n,line in enumerate(file):
117         if "Spin-dipolar (SD) contribution to J (Hz):" in line:
118             start = n
119             n = start + 1
120             resto = na%5
121             parts = int(na/5)
122             if (resto > 0):

```

```
123     parts += 1
124     for i in range(0, parts):
125         n = n + 1
126         for j in range(1, na - 5*i+1):
127
128             for k in range (0, min(j,5)):
129                 dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
130             n +=1
131     return (dic)
132
133 def pso_k(file):
134     dic = dict()
135     na = natoms(file)
136     for n,line in enumerate(file):
137         if "Paramagnetic spin-orbit (PSO) contribution to K (Hz):" in line:
138             start = n
139             n = start + 1
140             resto = na%5
141             parts = int(na/5)
142             if (resto > 0):
143                 parts += 1
144             for i in range(0, parts):
145                 n = n + 1
146                 for j in range(1, na - 5*i+1):
147
148                     for k in range (0, min(j,5)):
149                         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
150                     n +=1
151     return (dic)
152
153 def pso_j(file):
154     dic = dict()
155     na = natoms(file)
156     for n,line in enumerate(file):
157         if "Paramagnetic spin-orbit (PSO) contribution to J (Hz):" in line:
158             start = n
159             n = start + 1
160             resto = na%5
161             parts = int(na/5)
162             if (resto > 0):
163                 parts += 1
164             for i in range(0, parts):
165                 n = n + 1
166                 for j in range(1, na - 5*i+1):
167
168                     for k in range (0, min(j,5)):
169                         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
170                     n +=1
171     return (dic)
172
173 def dso_k(file):
174     dic = dict()
175     na = natoms(file)
176     for n,line in enumerate(file):
```

```

177         if "Diamagnetic spin-orbit (DSO) contribution to K (Hz):" in line:
178             start = n
179         n = start + 1
180         resto = na%5
181         parts = int(na/5)
182         if (resto > 0):
183             parts += 1
184         for i in range(0, parts):
185             n = n + 1
186             for j in range(1, na - 5*i+1):
187
188                 for k in range (0, min(j,5)):
189                     dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
190                 n +=1
191         return (dic)
192
193 def dso_j(file):
194     dic = dict()
195     na = natoms(file)
196     for n,line in enumerate(file):
197         if "Diamagnetic spin-orbit (DSO) contribution to J (Hz):" in line:
198             start = n
199         n = start + 1
200         resto = na%5
201         parts = int(na/5)
202         if (resto > 0):
203             parts += 1
204         for i in range(0, parts):
205             n = n + 1
206             for j in range(1, na - 5*i+1):
207
208                 for k in range (0, min(j,5)):
209                     dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
210                 n +=1
211         return (dic)
212
213 def tnss_k(file):
214     dic = dict()
215     na = natoms(file)
216     for n,line in enumerate(file):
217         if "Total nuclear spin-spin coupling K (Hz):" in line:
218             start = n
219         n = start + 1
220         resto = na%5
221         parts = int(na/5)
222         if (resto > 0):
223             parts += 1
224         for i in range(0, parts):
225             n = n + 1
226             for j in range(1, na - 5*i+1):
227
228                 for k in range (0, min(j,5)):
229                     dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
230                 n +=1

```



```
231     return (dic)
232
233 def tnss_j(file):
234     dic = dict()
235     na = natoms(file)
236     for n,line in enumerate(file):
237         if "Total nuclear spin-spin coupling J (Hz):" in line:
238             start = n
239             n = start + 1
240             resto = na%5
241             parts = int(na/5)
242             if (resto > 0):
243                 parts += 1
244             for i in range(0, parts):
245                 n = n + 1
246                 for j in range(1,na - 5*i+1):
247
248                     for k in range (0, min(j,5)):
249                         dic[(j+5*i,5*i+k+1)]=float((file[n].split()[k+1]).replace("D", "E"))
250                         n +=1
251     return (dic)
```

El archivo `elements.py` se encuentra en el árbol de directorios mostrado porque es empleado por el programa. Contiene información sobre propiedades de los elementos químicos y puede ser usado como un módulo. Fue desarrollado por Christoph Gohlke (<http://www.lfd.uci.edu/~gohlke/>).

Estudio conformacional

El primer paso para el estudio conformacional es la generación de los rotámeros (sección 8.2.4). El código desarrollado para dicha etapa se encuentra a continuación. Para su funcionamiento debe colocarse los archivos XYZ de las geometrías iniciales en el directorio principal de ejecución del programa. Este generará un subdirectorio por cada estructura con al menos un enlace rotable. Los mismos contendrán a los archivos XYZ de sus respectivos rotámeros.

```
1  # -*- coding: utf-8 -*-
2  import openbabel as ob
3  import pybel
4  from math import radians
5  import itertools, os, random
6
7  random.seed(666)
8
9  def gen_rot(name,nrot, **kwargs):
10     for mol in pybel.readfile( "xyz", name):
11         r1 = []
```

```

12 for bd in ob.OBMolBondIter(mol.OBMol):
13     if (bd.IsRotor()):
14         rl.append([bd.GetBeginAtomIdx(),bd.GetEndAtomIdx()])
15 if (len(rl) == 0):
16     print("No rotatable bonds")
17     return
18
19 lr14 = []
20 lt = []
21 for ln in rl:
22     at2ID = ln[0]
23     at3ID = ln[1]
24     at2 = mol.atoms[at2ID-1]
25     at3 = mol.atoms[at3ID-1]
26     for at2Vec in ob.OBAtomAtomIter(at2.OBAtom):
27         if (at2Vec.GetIdx() != at3ID):
28             at1ID = at2Vec.GetIdx()
29         for at3Vec in ob.OBAtomAtomIter(at3.OBAtom):
30             if (at3Vec.GetIdx() != at2ID):
31                 at4ID = at3Vec.GetIdx()
32             lr14.append([at1ID,at4ID])
33             lt.append(mol.OBMol.GetTorsion(at1ID,at2ID,at3ID,at4ID))
34
35 angs = []
36 for i in range(0,len(lt)):
37     angs.append([])
38     for n in range(0,nrot):
39         angs[i].append(radians(lt[i]+360./float(nrot)*float(n)))
40
41 if 'limit' in kwargs:
42     limit = kwargs['limit']
43 else:
44     limit = None
45
46 tabla = []
47 if limit is not None:
48     j = -1
49     for i in itertools.product(*angs):
50         j = j + 1
51         if j < limit:
52             tabla.append(list(i))
53         else:
54             k = random.randint(0, j)
55             if k < limit:
56                 tabla[k] = list(i)
57 else:
58     for i in itertools.product(*angs):
59         tabla.append(list(i))
60
61 for k in range(0,len(tabla)):
62     for l in range(0,len(tabla[k])):
63         at1 = mol.atoms[lr14[l][0]-1]
64         at2 = mol.atoms[rl[l][0]-1]
65         at3 = mol.atoms[rl[l][1]-1]

```

```

66         at4 = mol.atoms[lr14[1][1]-1]
67
68         mol.OBMol.SetTorsion(at1.OBAtom,at2.OBAtom,at3.OBAtom,at4.OBAtom,tabla[k][1])
69         print("Writing file for",str(k).zfill(4))
70         mol.write("xyz", str(k).zfill(4)+".xyz",overwrite=True)
71     return
72
73 dir_main = os.getcwd()
74 for i in os.listdir(dir_main):
75     if i.endswith(".xyz"):
76         inp_file = dir_main+"/"+i
77         base = i[:-4]
78         print("")
79         print("FILE")
80         print(base)
81         os.system("mkdir "+ base)
82         os.chdir(dir_main+"/"+base)
83         gen_rot(inp_file,3)
84         os.chdir(dir_main)

```

A los archivos así generados se los procesó luego con PM7 utilizando Mopac2012. El código siguiente analiza los resultados finales. Descarta las estructuras cuya conectividad sea diferente de la original. También detecta algunos casos de estructuras repetidas y elimina los duplicados. Para ello se implementó en el código el algoritmo de Kabsch. Este es un método para calcular la matriz de rotación óptima, en tanto que luego de su aplicación la distancia cuadrática media entre dos conjuntos apareados de puntos es mínima. El programa genera un archivo de salida con un informe de las energías de los distintos conformeros.

```

1  #- coding: utf-8 -*-
2  import sys
3  sys.path.append('/home/hernan/Bibliotecas/Python')
4  from elements.elements import ELEMENTS as elemento
5  import os
6  import numpy as np
7
8
9  def cm(molec):
10     M, x, y, z = 0., 0., 0., 0.
11     for atomo in molec:
12         mi = elemento[atomo[0]].mass
13         M = M + mi
14         x = x + mi * float(atomo[1])
15         y = y + mi * float(atomo[2])
16         z = z + mi * float(atomo[3])
17     return([x / M, y / M, z / M])
18
19 def trcm(molecula):
20     tr = cm(molecula)
21     mol = []

```

```
22     for atomo in molecula:
23         x = float(atomo[1]) - tr[0]
24         y = float(atomo[2]) - tr[1]
25         z = float(atomo[3]) - tr[2]
26         mol.append([atomo[0], x, y, z])
27     return(mol)
28
29 def only_xyz(molec):
30     natom = len(molec)
31     xyz = []
32     for i in range(natom):
33         xyz.append(
34             [float(molec[i][1]), float(molec[i][2]), float(molec[i][3])])
35     return(xyz)
36
37 def conn_atom(at1, at2, **kwargs):
38     if ('tol' in kwargs):
39         tol = float(kwargs['tol'])/100.
40     else:
41         tol = 0.2
42     dc = elemento[at1[0]].covrad + elemento[at2[0]].covrad
43     dr = d( only_xyz([at1]) , only_xyz([at2]) )
44     if (dr <= dc *(1.+tol)):
45         return(True)
46     else:
47         return(False)
48
49 def bonds_mol(mol, **kwargs):
50     if ('tol' in kwargs):
51         tol = kwargs['tol']
52     else:
53         tol = 20.
54
55     bonds = set()
56     for i in range(0,len(mol)-1):
57         for j in range(i+1,len(mol)):
58             if conn_atom(mol[i],mol[j],tol=tol):
59                 bonds.add((i,j))
60     return(bonds)
61
62 def listtoxyz(lista,archivo,**kwargs):
63     salida = open(archivo,"w")
64     print(len(lista),file=salida)
65     if 'comment' in kwargs:
66         print(kwargs['comment'],file=salida)
67     else:
68         print("",file=salida)
69     for at in lista:
70         print(at[0],at[1],at[2],at[3],file=salida)
71     return
72
73 def arc_coords(file):
74     lst = []
75     aux = False
```

```

76     for line in reversed(file):
77         if len(line.strip()) > 0:
78             lline = line.split()
79             lst.append([lline[0], lline[1], lline[3], lline[5]])
80             aux = True
81         else:
82             if (aux):
83                 lst.reverse()
84                 return(lst)
85
86 def arc_hof(file):
87     lst = []
88     aux = False
89     for line in file:
90         if "HEAT OF FORMATION" in line:
91             return(float(line.split()[4]))
92
93 def pairing(P,V):
94     Q = np.copy(V)
95     pair = np.zeros(P.shape[0],dtype=np.int)
96     for i in range(0,P.shape[0]):
97         p_i = np.tile(P[i,:],(P.shape[0],1))
98         normas = np.linalg.norm(p_i-Q, axis=1)
99         me_quedo_con = np.argmin(normas)
100        pair[i] = me_quedo_con
101        Q[me_quedo_con,:] = float("Inf")
102
103    return(P[pair])
104
105 np.set_printoptions(suppress=True)
106 np.set_printoptions(formatter={'float': '{: 0.1f}'.format})
107
108 def kabsch(P, Q):
109     A = (P.T).dot(Q)
110     V, S, W = np.linalg.svd(A)
111     I = np.identity(3)
112     I[2,2] = np.sign( np.linalg.det(V) * np.linalg.det(W) )
113     U = ((W.T).dot(I)).dot(V.T)
114     return(U)
115
116 def rmsd(A, B):
117     return( np.sqrt( (1./A.size) * np.sum((A-B)**2) ) )
118
119 def center(A):
120     return(A[:, :] - np.mean(A,axis=0) )
121
122 def kabsch_rmsd(P, Q):
123     U = kabsch(P, Q)
124     P = (U.dot(P.T)).T
125     for i in range(0,3):
126         P = pairing(P,Q)
127         U = kabsch(P, Q)
128         P = (U.dot(P.T)).T
129     return(rmsd(P, Q))

```

```
130
131 def d(p1, p2):
132     p = np.array(p1) - np.array(p2)
133     return(np.asscalar(np.sqrt((p).dot(p.T))))
134 def d_list(molec):
135     n = molec.shape[0]
136     lista = [ d(molec[i],molec[j]) for i in range(0, n-1) for j in range(i+1, n)]
137     return(np.array(lista))
138
139 dir_main = os.getcwd()
140
141 directorios = sorted([name for name in os.listdir(dir_main) if os.path.isdir(os.path.join(dir_main, name))])
142
143 os.makedirs(dir_main+"/REPORT")
144
145 archivos_analizados = 0
146 archivos_a_borrar = 0
147
148 energias_minimas = []
149
150 for i in directorios:
151     subdir = dir_main+"/"+i
152     print("Trabajando sobre el directorio", i)
153     os.chdir(dir_main+"/"+i)
154     arc_files = sorted([name for name in os.listdir(subdir) if name.endswith('arc') ])
155     mol_list = []
156     mol_enteras = []
157
158     for j in arc_files:
159         inp = open(j, "r")
160         f = inp.readlines()
161         inp.close()
162         mol_entera = arc_coords(f)
163         mol_entera = trcm(mol_entera)
164         mol = only_xyz(mol_entera)
165         mol_enteras.append(mol_entera[:])
166         mol_list.append(np.array(mol))
167
168     archivos_analizados += len(mol_list)
169     eliminar=[]
170     connect_right = bonds_mol(mol_enteras[0])
171
172     for k in range(1, len(mol_list)):
173         connect_k = bonds_mol(mol_enteras[k])
174         if ( connect_k != connect_right):
175             eliminar.append(k)
176
177     for k in range(0, len(mol_list)-1):
178         for j in range(k+1,len(mol_list)):
179             if ( k not in eliminar and j not in eliminar):
180                 iguales = False
181                 rmse = kabsch_rmsd(mol_list[k],mol_list[j])
182                 print(arc_files[k],arc_files[j],rmse)
183                 if (rmse < 0.1):
```

```
184         iguales = True
185         if (iguales):
186             eliminar.append(j)
187             print(k,j)
188
189     mantener = list( set(range(0,len(mol_list))) - set(eliminar) )
190
191     energy_array = np.zeros((len(mantener),2))
192     for k, j in enumerate(mantener):
193         energy_array[k,1] = arc_hof(open(arc_files[j]).readlines())
194         energy_array[k,0] = float(j)
195
196     minimo_tx = np.min(energy_array[:,1])
197     energy_array[:,1] = energy_array[:,1] - np.min(energy_array[:,1])
198     energias_minimas.append( (round(minimo_tx,2), i) )
199     energy_array = energy_array[energy_array[:,1].argsort()]
200     energy_file = open(dir_main+"/REPORT/"+i+"-summary.dat","w")
201     print("Standard enthalpies of formation in kcal/mol, PM7", file=energy_file)
202     print("Reference value: ", "%.2f" % round(minimo_tx,2), " kcal/mol", file=energy_file)
203
204     print("", file=energy_file)
205     for k, j in enumerate(energy_array[:,0]):
206         print("%.2f" % round(energy_array[k,1],2), i+"-"+str(int(j))+".xyz", file=energy_file)
207         listtoxyz(mol_enteras[int(j)], dir_main+"/REPORT/"+i+"-"+str(int(j))+".xyz",
208                 comment="Energía (PM7) respecto del rotámero más estable: "+str(energy_array[k,1])+
209                 "kcal/mol. HOF del más estable: "+str("%.2f" % round(minimo_tx,2))+" kcal/mol.")
210
211     file_minimos = open(dir_main+"/REPORT/minimum.dat","w")
212     for k in range(0, len(energias_minimas)):
213         print("%.2f" % round(energias_minimas[k][0],2), energias_minimas[k][1], file=file_minimos)
```


Bibliografía

- [1] Leonard Susskind and George Hrabovsky. *The theoretical minimum: what you need to know to start doing physics*. Basic Books, 2014.
- [2] Wikipedia. *Canonical quantization*, *The Free Encyclopedia*. 2017. Online; accessed 13-February-2017.
- [3] Albert Einstein. Zur elektrodynamik bewegter körper. *Ann. Phys.*, 322(10):891–921, 1905.
- [4] WA Von Ignatowsky. Einige allgemeine bemerkungen zum relativitätsprinzip. *Verh. Deutsch. Phys. Ges*, 12:788–796, 1910.
- [5] Palash B Pal. Nothing but relativity. *Eur. J. Phys.*, 24(3):315, 2003.
- [6] Joel W Gannett. Nothing but relativity, redux. *Eur. J. Phys.*, 28(6):1145, 2007.
- [7] Markus Reiher and Alexander Wolf. Exact decoupling of the Dirac Hamiltonian. I. General theory. *J. Chem. Phys.*, 121(5):2037–2047, 2004.
- [8] Leslie E Ballentine. *Quantum mechanics: a modern development*. World Scientific Publishing Co Inc, 2014.
- [9] David J Griffiths. *Introduction to electrodynamics, 3rd Edition*. Prentice Hall, 2007.
- [10] Herbert Goldstein. The classical motion of a rigid charged body in a magnetic field. *Am. J. Phys.*, 19(2):100–109, 1951.
- [11] Robert J Deissler. Dipole in a magnetic field, work, and quantum spin. *Phys. Rev. E*, 77(3):036609, 2008.
- [12] Antal Rockenbauer. Can the spinning of elementary particles produce the rest energy mc^2 the vortex model of elementary particles. *arXiv preprint arXiv:0808.1656*, 2008.

- [13] Trygve Helgaker, Michał Jaszuński, and Kenneth Ruud. Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants. *Chem. Rev.*, 99(1):293–352, 1999.
- [14] Trygve Helgaker and Poul Jorgensen. An electronic hamiltonian for origin independent calculations of magnetic properties. *J. Chem. Phys.*, 95(4):2595–2601, 1991.
- [15] Werner Kutzelnigg. Ab initio calculation of molecular properties. *J. Mol. Struct. THEOCHEM*, 202:11–61, 1989.
- [16] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.
- [17] Elliott H Lieb. Density functionals for coulomb-systems. *Int. J. Quantum Chem.*, 24(3):243–277, 1983.
- [18] Llewellyn H Thomas. The calculation of atomic fields. In *Math. Proc. Cambridge*, volume 23, pages 542–548. Cambridge Univ Press, 1927.
- [19] Enrico Fermi. Un metodo statistico per la determinazione di alcune priorietà dell'atome. *Rend. Accad. Naz. Lincei*, 6(602-607):32, 1927.
- [20] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140(4A):A1133, 1965.
- [21] Seymour H Vosko, Leslie Wilk, and Marwan Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.
- [22] John P Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23(10):5048, 1981.
- [23] Axel D. Becke. Density functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652, 1993.
- [24] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, Jan 1988.

- [25] John P Perdew, Matthias Ernzerhof, and Kieron Burke. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, 105(22):9982–9985, 1996.
- [26] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *J. Chem. Phys.*, 110(13):6158–6170, 1999.
- [27] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211, 1980.
- [28] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627, 1994.
- [29] Vladimir Sychrovský, Jürgen Gräfenstein, and Dieter Cremer. Nuclear magnetic resonance spin–spin coupling constants from coupled perturbed density functional theory. *J. Chem. Phys.*, 113(9):3530–3547, 2000.
- [30] James R Cheeseman, Gary W Trucks, Todd A Keith, and Michael J Frisch. A comparison of models for calculating nuclear magnetic resonance shielding tensors. *J. Chem. Phys.*, 104(14):5497–5509, 1996.
- [31] Erik I Tellgren, Trygve Helgaker, and Alessandro Soncini. Non-perturbative magnetic phenomena in closed-shell paramagnetic molecules. *Phys. Chem. Chem. Phys.*, 11(26):5489–5498, 2009.
- [32] Erik I Tellgren, Simen Kvaal, Espen Sagvolden, Ulf Ekström, Andrew M Teale, and Trygve Helgaker. Choice of basic variables in current-density-functional theory. *Phys. Rev. A*, 86(6):062506, 2012.
- [33] Andre Laestadius and Michael Benedicks. Hohenberg–kohn theorems in the presence of magnetic field. *Int. J. Quantum Chem.*, 114(12):782–795, 2014.
- [34] Simen Kvaal and Trygve Helgaker. Ground-state densities from the rayleigh–ritz variation principle and from density-functional theory. *J. Chem. Phys.*, 143(18):184106, 2015.

- [35] Virahat Sahni. Generalized Hohenberg-Kohn Theorems in Electrostatic and Magnetostatic Fields. In *Quantal Density Functional Theory*, pages 253–282. Springer, 2016.
- [36] Sarah Reimann, Ulf Ekström, Stella Stopkowicz, Andrew M Teale, Alex Borgoo, and Trygve Helgaker. The importance of current contributions to shielding constants in density-functional theory. *Phys. Chem. Chem. Phys.*, 17(28):18834–18842, 2015.
- [37] Thomas W Keal and David J Tozer. The exchange-correlation potential in Kohn-Sham nuclear magnetic resonance shielding calculations. *J. Chem. Phys.*, 119(6):3015–3024, 2003.
- [38] Erik I Tellgren, Alessandro Soncini, and Trygve Helgaker. Nonperturbative ab initio calculations in strong magnetic fields using london orbitals. *J. Chem. Phys.*, 129(15):154114, 2008.
- [39] José M Bernardo and Adrian FM Smith. Bayesian theory, 2001.
- [40] Arthur M Doweyko. QSAR: dead or alive? *J. Comput. Aid. Mol. Des.*, 22(2):81–89, 2008.
- [41] Stephen R Johnson. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.*, 48(1):25–26, 2008.
- [42] Gregory C. Reinsel and Raja P. Velu. *Multivariate Linear Regression*, pages 1–14. Springer New York, New York, NY, 1998.
- [43] Pablo R Duchowicz, Eduardo A Castro, and Francisco M Fernández. Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. *MATCH Commun. Math. Comput. Chem*, 55:179–192, 2006.
- [44] Andrew G Mercader, Pablo R Duchowicz, Francisco M Fernández, and Eduardo A Castro. Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories. *Chemometr. Intell. Lab.*, 92(2):138–144, 2008.
- [45] Andrew G Mercader, Pablo R Duchowicz, Francisco M Fernández, and Eduardo A Castro. Advances in the replacement and enhanced replacement method in QSAR and QSPR theories. *J. Chem. Inf. Model.*, 51(7):1575–1581, 2011.

- [46] Andrew G Mercader and Pablo R Duchowicz. Enhanced replacement method integration with genetic algorithms populations in QSAR and QSPR theories. *Chemometr. Intell. Lab.*, 149:117–122, 2015.
- [47] Douglas M Hawkins, Subhash C Basak, and Denise Mills. Assessing model fit by cross-validation. *J. Chem. Inf. Comp. Sci.*, 43(2):579–586, 2003.
- [48] Alexander Golbraikh and Alexander Tropsha. Beware of q^2 ! *J. Mol. Graph. Model.*, 20(4):269–276, 2002.
- [49] Christoph Rücker, Gerta Rücker, and Markus Meringer. y -Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.*, 47(6):2345–2357, 2007.
- [50] Lennart Eriksson, Joanna Jaworska, Andrew P Worth, Mark TD Cronin, Robert M McDowell, and Paola Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives*, 111(10):1361, 2003.
- [51] Alexander Tropsha, Paola Gramatica, and Vijay K Gombar. The importance of being earnest: validation is the absolute essential for successful application and interpretation of qspr models. *Molecular Informatics*, 22(1):69–77, 2003.
- [52] Andrew Gustavo Mercader. *Desarrollo y aplicación de la teoría QSAR/QSPR*. PhD thesis, Facultad de Ciencias Exactas, 2009.
- [53] Laura K Schnackenberg and Richard D Beger. Whole-molecule calculation of log P based on molar volume, hydrogen bonds, and simulated ^{13}C NMR spectra. *J. Chem. Inf. Model.*, 45(2):360–365, 2005.
- [54] Igor V Tetko, Johann Gasteiger, Roberto Todeschini, Andrea Mauri, David Livingstone, Peter Ertl, Vladimir A Palyulin, Eugene V Radchenko, Nikolay S Zefirov, Alexander S Makarenko, et al. Virtual computational chemistry laboratory—design and description. *J. Comput. Aid. Mol. Des.*, 19(6):453–463, 2005.
- [55] Eric S Goll and Peter C Jurs. Prediction of the normal boiling points of organic compounds from molecular structures with a computational neural network model. *J. Chem. Inf. Comp. Sci.*, 39(6):974–983, 1999.

- [56] Aixia Yan and Johann Gasteiger. Prediction of aqueous solubility of organic compounds based on a 3d structure representation. *J. Chem. Inf. Comp. Sci.*, 43(2):429–434, 2003.
- [57] Python Software Foundation. Python language reference, version 3.6. available at <http://www.python.org>, 2017.
- [58] Peter J Linstrom and WG Mallard. Nist chemistry webbook; nist standard reference database no. 69. 2001.
- [59] Evan E Bolton, Yanli Wang, Paul A Thiessen, and Stephen H Bryant. Pubchem: integrated platform of small molecules and biological activities. *Annual Rep. Comput. Chem.*, 4:217–241, 2008.
- [60] Jiří Hostaš, Jan Řezáč, and Pavel Hobza. On the performance of the semiempirical quantum mechanical pm6 and pm7 methods for noncovalent interactions. *Chem. Phys. Lett.*, 568:161–166, 2013.
- [61] Stewart JJP MOPAC. Stewart computational chemistry, 2012.
- [62] Noel M O’Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *J. Cheminformatics*, 3(1):1, 2011.
- [63] Noel M O’Boyle, Chris Morley, and Geoffrey R Hutchison. Pybel: a python wrapper for the openbabel cheminformatics toolkit. *Chem. Cent. J.*, 2(1):5, 2008.
- [64] Ronald Aylmer Fisher and Frank Yates. *Statistical tables for biological, agricultural and medical research*. Longman, 1938.
- [65] Richard Durstenfeld. Algorithm 235: random permutation. *Commun. ACM*, 7(7):420, 1964.
- [66] Dirk Porezag, Th Frauenheim, Th Köhler, Gotthard Seifert, and R Kaschner. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B*, 51(19):12947, 1995.
- [67] G Seifert, D Porezag, and Th Frauenheim. Calculations of molecules, clusters, and solids with a simplified lcao-dft-lda scheme. *Int. J. Quantum Chem.*, 58(2):185–192, 1996.

- [68] Marcus Elstner, Dirk Porezag, G Jungnickel, J Elsner, M Haugk, Th Frauenheim, Sandor Suhai, and Gotthard Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58(11):7260, 1998.
- [69] Fritz London. Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J. Phys. Radium*, 8(10):397–409, 1937.
- [70] Rev McWeeny. Perturbation theory for the fock-dirac density matrix. *Phys. Rev.*, 126(3):1028, 1962.
- [71] Robert Ditchfield. Self-consistent perturbation theory of diamagnetism: I. A gauge-invariant LCAO method for NMR chemical shifts. *Mol. Phys.*, 27(4):789–807, 1974.
- [72] Krzysztof Wolinski, James F Hinton, and Peter Pulay. Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.*, 112(23):8251–8260, 1990.
- [73] Roberto Todeschini and Viviana Consonni. *Handbook of molecular descriptors*, volume 11. John Wiley & Sons, 2008.
- [74] Kunal Roy. On some aspects of validation of predictive quantitative structure–activity relationship models. *Expert Opin. Drug Dis.*, 2(12):1567–1577, 2007.
- [75] Veerasamy Ravichandran, Sivadasan Shalini, Karupiah Sundram, and Arumugam Dhanaraj Sokkalingam. QSAR study of substituted 1, 3, 4-oxadiazole naphthyridines as HIV-1 integrase inhibitors. *European journal of medicinal chemistry*, 45(7):2791–2797, 2010.
- [76] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.*, 13(2):22–30, 2011.